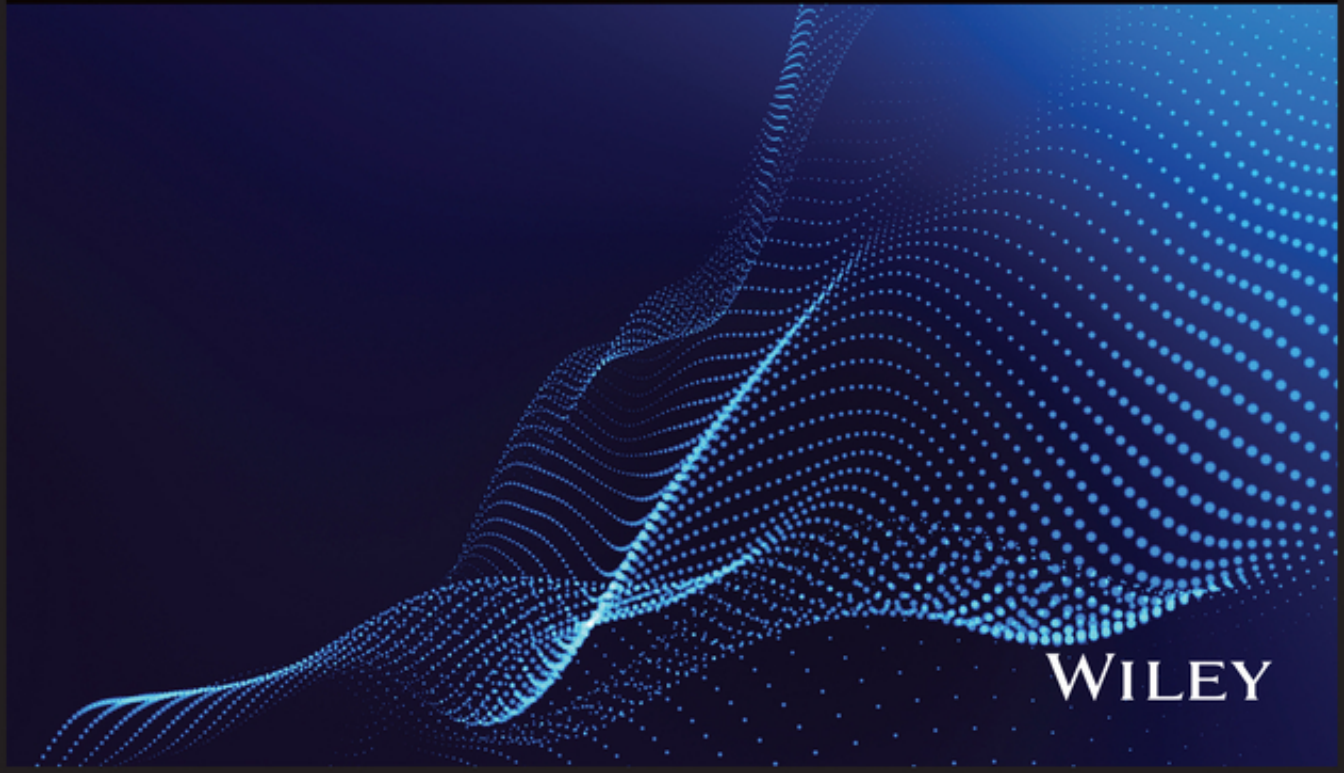


Wiley Finance Series

SIXTH EDITION

RISK MANAGEMENT AND FINANCIAL INSTITUTIONS

JOHN C. HULL



WILEY

Risk Management and Financial Institutions

Founded in 1807, John Wiley & Sons is the oldest independent publishing company in the United States. With offices in North America, Europe, Australia, and Asia, Wiley is globally committed to developing and marketing print and electronic products and services for our customers' professional and personal knowledge and understanding.

The Wiley Finance series contains books written specifically for finance and investment professionals as well as sophisticated individual investors and their financial advisors. Book topics range from portfolio management to e-commerce, risk management, financial engineering, valuation and financial instrument analysis, as well as much more.

For a list of available titles, visit our Web site at www.WileyFinance.com.

Risk Management and Financial Institutions

Sixth Edition

John C. Hull

WILEY

Copyright © 2023 by John C. Hull. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data is Available:

ISBN 9781119932482 (Hardback)

ISBN 9781119932505 (ePDF)

ISBN 9781119932499 (epub)

Cover Design: Wiley

Cover Image: © in-future/Getty Images

To Michelle, Peter, and David

Contents in Brief

Business Snapshots	xxiii
Preface	xxv
Chapter 1 Introduction: Risk-Return Trade-offs	1
Part 1: Financial Institutions	23
Chapter 2 Banks	25
Chapter 3 Insurance Companies and Pension Plans	47
Chapter 4 Fund Managers	75
Part 2: Financial Markets	97
Chapter 5 Financial Instruments	99
Chapter 6 The OTC Derivatives Market	129
Chapter 7 Securitization and the Global Financial Crisis	145
Chapter 8 Volatility	163
Chapter 9 Correlations and Copulas	193
Chapter 10 Valuation and Scenario Analysis	217

Part 3: Market Risk	231
Chapter 11 Value at Risk and Expected Shortfall	233
Chapter 12 Historical Simulation and Extreme Value Theory	257
Chapter 13 Model-Building Approach	279
Chapter 14 Interest Rate Risk	293
Chapter 15 Derivatives Risk	319
Chapter 16 Scenario Analysis and Stress Testing	347
Part 4: Credit Risk	365
Chapter 17 Estimating Default Probabilities	367
Chapter 18 xVAs	393
Chapter 19 Credit Value at Risk	413
Part 5: Other Risks	429
Chapter 20 Operational Risk	431
Chapter 21 Liquidity Risk	449
Chapter 22 Model Risk Management	477
Chapter 23 Climate Risk, ESG, and Sustainability	497
Chapter 24 Enterprise Risk Management	513
Part 6: Regulation	531
Chapter 25 Basel I, Basel II, and Solvency II	533
Chapter 26 Basel II.5, Basel III, and Other Post-Crisis Changes	563
Chapter 27 Fundamental Review of the Trading Book	585
Chapter 28 Economic Capital and RAROC	599

Part 7: Other Topics	617
Chapter 29 Financial Innovation	619
Chapter 30 Risk Management Mistakes to Avoid	641
Part 8: Appendices	653
Appendix A Compounding Frequencies for Interest Rates	655
Appendix B Zero Rates, Forward Rates, and Zero-Coupon Yield Curves	659
Appendix C Valuing Forward and Futures Contracts	663
Appendix D Valuing Swaps	665
Appendix E Valuing European Options	669
Appendix F Valuing American Options	673
Appendix G Taylor Series Expansions	677
Appendix H Eigenvectors and Eigenvalues	681
Appendix I Principal Components Analysis	685
Appendix J Manipulation of Credit Transition Matrices	687
Appendix K Valuation of Credit Default Swaps	689
Appendix L Synthetic CDOs and Their Valuation	693
Appendix M SIMM	697
Answers to Questions and Problems	701
Glossary	743
RMFI Software	771
Table for $N(x)$ When $x \geq 0$	775
Table for $N(x)$ When $x \leq 0$	777
Index	779

Contents

Business Snapshots	xxiii
Preface	xxv
Chapter 1	Introduction: Risk-Return Trade-offs
1.1	Risk vs. Return for Investors
1.2	The Efficient Frontier
1.3	The Capital Asset Pricing Model
1.4	Arbitrage Pricing Theory
1.5	Risk vs. Return for Companies
1.6	Risk Management by Financial Institutions
1.7	Credit Ratings
	Summary
	Further Reading
	Practice Questions and Problems (Answers at End of Book)
	Further Questions
	Part 1: Financial Institutions
Chapter 2	Banks
2.1	Commercial Banking
2.2	The Capital Requirements of a Small Commercial Bank
2.3	Deposit Insurance
2.4	Investment Banking

2.5	Securities Trading	37
2.6	Potential Conflicts of Interest in Banking	38
2.7	Today's Large Banks	39
2.8	The Risks Facing Banks	42
	Summary	43
	Further Reading	43
	Practice Questions and Problems (Answers at	
	End of Book)	43
	Further Questions	44
Chapter 3	Insurance Companies and Pension Plans	47
3.1	Life Insurance	48
3.2	Annuity Contracts	51
3.3	Mortality Tables	53
3.4	Longevity and Mortality Risk	56
3.5	Property-Casualty Insurance	57
3.6	Health Insurance	60
3.7	Moral Hazard and Adverse Selection	61
3.8	Reinsurance	62
3.9	Capital Requirements	63
3.10	The Risks Facing Insurance Companies	64
3.11	Regulation	65
3.12	Pension Plans	66
	Summary	70
	Further Reading	71
	Practice Questions and Problems (Answers at	
	End of Book)	71
	Further Questions	72
Chapter 4	Fund Managers	75
4.1	Mutual Funds	75
4.2	Exchange-Traded Funds	79
4.3	Active vs. Passive Management	80
4.4	Regulation	83
4.5	Hedge Funds	84
4.6	Hedge Fund Strategies	89
4.7	Hedge Fund Performance	93
	Summary	94
	Further Reading	95
	Practice Questions and Problems (Answers at	
	End of Book)	95
	Further Questions	96

Part 2: Financial Markets		97
Chapter 5	Financial Instruments	99
5.1	Long and Short Positions in Assets	99
5.2	Derivatives Markets	102
5.3	Plain Vanilla Derivatives	103
5.4	Non-Traditional Derivatives	115
5.5	Exotic Options and Structured Products	119
5.6	Risk Management Challenges	123
	Summary	123
	Further Reading	123
	Practice Questions and Problems (Answers at End of Book)	124
	Further Questions	126
Chapter 6	The OTC Derivatives Market	129
6.1	A Reference Point: Exchange-Traded Markets	129
6.2	Clearing in OTC Derivatives Markets	131
6.3	Post-Crisis Regulatory Changes	136
6.4	Impact of the Changes	138
6.5	CCPs and Bankruptcy	141
	Summary	142
	Further Reading	142
	Practice Questions and Problems (Answers at End of Book)	143
	Further Questions	143
Chapter 7	Securitization and the Global Financial Crisis	145
7.1	The U.S. Housing Market	146
7.2	Securitization	149
7.3	The Losses	155
7.4	What Went Wrong?	156
7.5	Lessons from the Global Financial Crisis	158
	Summary	159
	Further Reading	160
	Practice Questions and Problems (Answers at End of Book)	160
	Further Questions	161
Chapter 8	Volatility	163
8.1	Definition of Volatility	163
8.2	Implied Volatilities	165
8.3	Are Daily Percentage Changes in Financial Variables Normal?	167

8.4	The Power Law	170
8.5	Monitoring Daily Volatility	172
8.6	The Exponentially Weighted Moving Average Model	175
8.7	The GARCH(1,1) Model	177
8.8	Choosing between the Models	179
8.9	Maximum Likelihood Methods	179
8.10	Using GARCH(1,1) to Forecast Future Volatility	185
	Summary	189
	Further Reading	189
	Practice Questions and Problems (Answers at End of Book)	190
	Further Questions	191
Chapter 9	Correlations and Copulas	193
9.1	Definition of Correlation	193
9.2	Monitoring Correlation	195
9.3	Correlation and Covariance Matrices	198
9.4	Multivariate Normal Distributions	199
9.5	Copulas	202
9.6	Application to Loan Portfolios: Vasicek's Model	208
	Summary	213
	Further Reading	213
	Practice Questions and Problems (Answers at End of Book)	214
	Further Questions	215
Chapter 10	Valuation and Scenario Analysis	217
10.1	Volatility and Asset Prices	218
10.2	Risk-Neutral Valuation	219
10.3	Scenario Analysis	224
10.4	When Both Worlds Have to Be Used	224
10.5	The Calculations in Practice	225
	Summary	227
	Further Reading	228
	Practice Questions and Problems (Answers at End of Book)	228
	Further Questions	229
	Part 3: Market Risk	231
Chapter 11	Value at Risk and Expected Shortfall	233
11.1	Definition of VaR	233
11.2	Examples of the Calculation of VaR	235
11.3	A Drawback of VaR	237

11.4	Expected Shortfall	238
11.5	Coherent Risk Measures	238
11.6	Choice of Parameters for VaR and ES	242
11.7	Marginal, Incremental, and Component Measures	247
11.8	Euler's Theorem	248
11.9	Aggregating VaRs and ESs	249
11.10	Back-Testing	249
	Summary	253
	Further Reading	253
	Practice Questions and Problems (Answers at End of Book)	254
	Further Questions	255
Chapter 12	Historical Simulation and Extreme Value Theory	257
12.1	The Methodology	257
12.2	Accuracy of VaR	263
12.3	Extensions	265
12.4	Computational Issues	270
12.5	Extreme Value Theory	270
12.6	Applications of EVT	273
	Summary	276
	Further Reading	276
	Practice Questions and Problems (Answers at End of Book)	277
	Further Questions	278
Chapter 13	Model-Building Approach	279
13.1	The Basic Methodology	280
13.2	Generalization	282
13.3	The Four-Index Example Revisited	284
13.4	Extensions of the Basic Procedure	286
13.5	Risk Weights and Weighted Sensitivities	287
13.6	Non-Linearity	288
13.7	Model-Building vs. Historical Simulation	288
	Summary	289
	Further Reading	289
	Practice Questions and Problems (Answers at End of Book)	289
	Further Questions	290
Chapter 14	Interest Rate Risk	293
14.1	Types of Rates	294
14.2	Calculating Risk Measures	296

14.3	Principal Components Analysis	299
14.4	The Management of Net Interest Income	303
14.5	Duration	305
14.6	Convexity	308
14.7	Generalization	310
14.8	Nonparallel Yield Curve Shifts	312
	Summary	315
	Further Reading	316
	Practice Questions and Problems (Answers at End of Book)	316
	Further Questions	317
Chapter 15	Derivatives Risk	319
15.1	Delta	319
15.2	Gamma	327
15.3	Vega	329
15.4	Theta	331
15.5	Rho	332
15.6	Calculating Greek Letters	332
15.7	Taylor Series Expansions	333
15.8	The Realities of Hedging Derivatives	335
15.9	Hedging Exotic Options	336
15.10	Scenario Analysis	338
15.11	Approximate Analytical Results	339
	Summary	343
	Further Reading	344
	Practice Questions and Problems (Answers at End of Book)	344
	Further Questions	345
Chapter 16	Scenario Analysis and Stress Testing	347
16.1	Generating the Scenarios	347
16.2	Regulation	354
16.3	What to Do with the Results	358
	Summary	361
	Further Reading	361
	Practice Questions and Problems (Answers at End of Book)	362
	Further Questions	363
	Part 4: Credit Risk	365
Chapter 17	Estimating Default Probabilities	367
17.1	Credit Ratings	367
17.2	Historical Default Probabilities	369

17.3	Recovery Rates	371
17.4	Credit Default Swaps	372
17.5	Credit Spreads	377
17.6	Estimating Default Probabilities from Credit Spreads	380
17.7	Comparison of Default Probability Estimates	382
17.8	Using Equity Prices to Estimate Default Probabilities	386
	Summary	388
	Further Reading	389
	Practice Questions and Problems (Answers at End of Book)	390
	Further Questions	391
Chapter 18	xVAs	393
18.1	Credit Exposure on Derivatives	394
18.2	CVA	395
18.3	The Impact of a New Transaction	399
18.4	CVA Risk	400
18.5	Wrong-Way Risk	401
18.6	DVA	402
18.7	Some Simple Examples	403
18.8	Other xVAs	407
	Summary	408
	Further Reading	409
	Practice Questions and Problems (Answers at End of Book)	409
	Further Questions	410
Chapter 19	Credit Value at Risk	413
19.1	Ratings Transition Matrices	414
19.2	Vasicek's Model	416
19.3	Credit Risk Plus	417
19.4	Creditmetrics	420
19.5	Credit Spread Risk	423
	Summary	425
	Further Reading	426
	Practice Questions and Problems (Answers at End of Book)	426
	Further Questions	427
	Part 5: Other Risks	429
Chapter 20	Operational Risk	431
20.1	Defining Operational Risk	433
20.2	Types of Operational Risk	434
20.3	Loss Severity and Loss Frequency	434

20.4	The Standardized Measurement Approach	439
20.5	Preventing Operational Risk Losses	441
20.6	Allocation of Operational Risk Capital	442
20.7	Use of Power Law	443
20.8	Insurance	443
20.9	Sarbanes–Oxley	445
	Summary	446
	Further Reading	446
	Practice Questions and Problems (Answers at End of Book)	447
	Further Questions	448
Chapter 21	Liquidity Risk	449
21.1	Liquidity Trading Risk	450
21.2	Liquidity Funding Risk	457
21.3	Liquidity Black Holes	466
	Summary	472
	Further Reading	473
	Practice Questions and Problems (Answers at End of Book)	474
	Further Questions	475
Chapter 22	Model Risk Management	477
22.1	Regulatory Guidance	478
22.2	Models in Physics and Finance	484
22.3	Simple Models: Expensive Mistakes	484
22.4	Models for Pricing Actively Traded Products	487
22.5	Models for Less Actively Traded Products	490
22.6	Accounting	492
22.7	What Makes a Successful Pricing Model?	492
22.8	Model-Building Missteps	493
	Summary	494
	Further Reading	495
	Practice Questions and Problems (Answers at End of Book)	495
	Further Questions	496
Chapter 23	Climate Risk, ESG, and Sustainability	497
23.1	Climate Risk	497
23.2	ESG	507
23.3	Sustainability	508
23.4	Greenwashing	509
	Summary	510
	Further Reading	510

	Practice Questions and Problems (Answers at End of Book)	511
	Further Questions	511
Chapter 24	Enterprise Risk Management	513
	24.1 Risk Appetite	514
	24.2 Risk Culture	520
	24.3 Identifying Major Risks	523
	24.4 Strategic Risk Management	526
	Summary	527
	Further Reading	528
	Practice Questions and Problems (Answers at End of Book)	528
	Further Questions	529
	Part 6: Regulation	531
Chapter 25	Basel I, Basel II, and Solvency II	533
	25.1 The Reasons for Regulating Banks	533
	25.2 Bank Regulation Pre-1988	535
	25.3 The 1988 BIS Accord	536
	25.4 The G-30 Policy Recommendations	539
	25.5 Netting	540
	25.6 The 1996 Amendment	542
	25.7 Basel II	545
	25.8 Credit Risk Capital Under Basel II	546
	25.9 Operational Risk Capital Under Basel II	555
	25.10 Pillar 2: Supervisory Review	556
	25.11 Pillar 3: Market Discipline	556
	25.12 Solvency II	557
	Summary	558
	Further Reading	559
	Practice Questions and Problems (Answers at End of Book)	559
	Further Questions	561
Chapter 26	Basel II.5, Basel III, and Other Post-Crisis Changes	563
	26.1 Basel II.5	564
	26.2 Basel III	567
	26.3 Contingent Convertible Bonds	576
	26.4 Use of Standardized Approaches and SA-CCR	576
	26.5 Dodd-Frank Act	578
	26.6 Legislation in Other Countries	580
	Summary	582
	Further Reading	582

	Practice Questions and Problems (Answers at End of Book)	583
	Further Questions	583
Chapter 27	Fundamental Review of the Trading Book	585
27.1	Background	586
27.2	Standardized Approach	588
27.3	Internal Models Approach	591
27.4	Trading Book vs. Banking Book	595
	Summary	596
	Further Reading	596
	Practice Questions and Problems (Answers at End of Book)	596
	Further Question	597
Chapter 28	Economic Capital and RAROC	599
28.1	Definition of Economic Capital	600
28.2	Components of Economic Capital	602
28.3	Shapes of the Loss Distributions	604
28.4	Relative Importance of Risks	605
28.5	Aggregating Economic Capital	606
28.6	Allocation of Economic Capital	610
28.7	Deutsche Bank's Economic Capital	611
28.8	RAROC	612
	Summary	613
	Further Reading	614
	Practice Questions and Problems (Answers at End of Book)	614
	Further Questions	615
	Part 7: Other Topics	617
Chapter 29	Financial Innovation	619
29.1	Technological Advances	620
29.2	Payment Systems	624
29.3	Open Banking	628
29.4	Lending	629
29.5	Wealth Management	631
29.6	InsurTech	632
29.7	Regulation and Compliance	634
29.8	How Should Financial Institutions Respond?	636
	Summary	639
	Further Reading	640

Practice Questions and Problems (Answers at End of Book)	640
Further Questions	640
Chapter 30 Risk Management Mistakes to Avoid	641
30.1 Risk Limits	643
30.2 Managing the Trading Room	645
30.3 Liquidity Risk	647
30.4 Lessons for Nonfinancial Corporations	650
30.5 A Final Point	651
Further Reading	652
Part 8: Appendices	653
Appendix A Compounding Frequencies for Interest Rates	655
Appendix B Zero Rates, Forward Rates, and Zero-Coupon Yield Curves	659
Appendix C Valuing Forward and Futures Contracts	663
Appendix D Valuing Swaps	665
Appendix E Valuing European Options	669
Appendix F Valuing American Options	673
Appendix G Taylor Series Expansions	677
Appendix H Eigenvectors and Eigenvalues	681
Appendix I Principal Components Analysis	685
Appendix J Manipulation of Credit Transition Matrices	687
Appendix K Valuation of Credit Default Swaps	689
Appendix L Synthetic CDOs and Their Valuation	693
Appendix M SIMM	697
Answers to Questions and Problems	701
Glossary	743
RMFI Software	771
Table for $N(x)$ When $x \geq 0$	775
Table for $N(x)$ When $x \leq 0$	777
Index	779

Business Snapshots

1.1	The Hidden Costs of Bankruptcy: A Hypothetical Scenario	16
2.1	Google's IPO	35
2.2	Donald Trump and a SPAC	36
2.3	How to Avoid Loan Losses	41
3.1	Equitable Life	53
3.2	A Perfect Storm	68
4.1	Mutual Fund Returns Can Be Misleading	82
5.1	The Unanticipated Delivery of a Futures Contract	106
5.2	A Software Error?	107
5.3	Microsoft's Hedging	120
5.4	Procter and Gamble's Bizarre Deal	121
5.5	SocGen's Big Loss in 2008	122
6.1	Rehypothecation	140
7.1	All BBBs Are Not the Same	155
7.2	A Trading Opportunity?	159
8.1	What Causes Volatility?	166
8.2	Making Money from Foreign Currency Options	169
11.1	Historical Perspectives on VaR	234
15.1	Hedging by Gold Mining Companies	322
15.2	Dynamic Hedging in Practice	336
15.3	Is Delta Hedging Easier or More Difficult for Exotics?	337
16.1	Long-Term Capital Management's Big Loss	351
16.2	Traffic Light Options	357
17.1	The CDS Market	373
17.2	Is a CDS a Flawed Contract?	375

17.3	Contagion	385
20.1	The Hammersmith and Fulham Story	440
20.2	Rogue Trader Insurance	444
21.1	Northern Rock	458
21.2	Ashanti Goldfields	462
21.3	Metallgesellschaft	463
21.4	The Crash of 1987	468
22.1	The London Whale	481
22.2	Kidder Peabody's Embarrassing Mistake	485
22.3	Exploiting the Weaknesses of a Competitor's Mode	486
23.1	Transition Costs in Germany	501
23.2	Reputational Risk and HSBC	505
24.1	Cross-Selling at Wells Fargo	518
24.2	The Santander Rail Deal	521
24.3	ABACUS	522
25.1	Systemic risk	534
26.1	Credit Suisse's CoCo Bond Issues	577
28.1	The EGT Fund	606
29.1	Mt. Gox	627
29.2	Eastman Kodak and the Digitization of Photography	637
30.1	Big Losses	642

Preface

The sixth edition of *Risk Management and Financial Institutions* has been fully updated with many improvements to the presentation of material. Like my other popular book, *Options, Futures and Other Derivatives*, it is designed to be suitable for practicing managers as well as university students. Those studying for professional qualifications such as FRM and PRM will also find the book useful.

There are no prerequisites, except a knowledge of basic statistics. In producing the sixth edition I re-sequenced the chapters to create a more logical flow of material. Financial institutions, financial markets, and the management of different types of risk are now discussed in the first 24 chapters, with the regulation of financial institutions being covered toward the end of the book. Instructors who do not want to cover regulation will find the new structure easier to deal with. Those who do cover regulation will probably agree that students should have a good understanding of the risks first.

There is a new chapter (Chapter 23) on climate risk, ESG, and sustainability. We have now reached the stage where all companies, including financial institutions, cannot ignore environmental risks. Increasingly, it is necessary for them to provide information to investors on the way they are managing these risks. This chapter describes the nature of environmental risks and the key role that financial institutions can play in saving the planet.

Improvements have been made to material throughout the book. The level of mathematical sophistication and the way material is presented have been managed carefully so that the book is accessible to as wide an audience as possible. For example, when covering copulas in Chapter 9, I present the intuition followed by a detailed numerical example; when covering maximum likelihood methods in Chapter 8 and extreme value

theory in Chapter 12, I provide illustrative numerical examples. I have also provided Excel spreadsheets for many of the applications in the book on my website:

www-2.rotman.utoronto.ca/~hull

Ancillary Material

Several hundred PowerPoint slides can be downloaded from my website or from the Wiley Higher Education website. Adopting instructors are welcome to adapt the slides to meet their own needs. The RMFI software that accompanies the book is also available from my website.

End-of-chapter problems are divided into two groups: “Practice Questions and Problems” and “Further Questions.” Answers to the former are at the end of the book. Answers to the latter and accompanying worksheets are available to adopting instructors from the Wiley Higher Education website. This website also contains suggestions on the teaching of each chapter and course organization.

Acknowledgments

I have benefited from interactions with many academics and practicing risk managers when writing this book. I would particularly like to thank Jan Mahrt-Smith (Rotman), Jun Yuan (Royal Bank of Canada), and Beth Creller (GARP) who made suggestions on how Chapter 23 could be improved. I would also like to thank students in the MBA, Master of Finance, and Master of Financial Risk Management courses at the Rotman School, University of Toronto, who have made many useful suggestions.

Alan White, a long-time colleague at Rotman, deserves a special acknowledgment. Alan and I have been carrying out joint research and consulting for many years. We have spent countless hours discussing key issues. Many of the new ideas in this book, and many of the new ways to explain old ideas, are as much Alan’s as mine. Alan has done most of the development work on the RMFI software.

Special thanks are due to many people at Wiley, particularly Bill Falloon, Purvi Patel, Samantha Wu, Susan Cerra, and Premkumar for their enthusiasm, advice, and encouragement.

I welcome comments on the book from readers. My e-mail address is:

hull@rotman.utoronto.ca

JOHN HULL
*Joseph L Rotman School of Management
University of Toronto*

Chapter 1

Introduction: Risk-Return Trade-offs

Imagine you are the Chief Risk Officer (CRO) of a major corporation. The Chief Executive Officer (CEO) wants your views on a major new venture. You have been inundated with reports showing that the new venture has a positive net present value and will enhance shareholder value. What sort of analysis and ideas is the CEO looking for from you?

As CRO it is your job to consider how the new venture fits into the company's portfolio. What is the correlation of the performance of the new venture with the rest of the company's business? When the rest of the business is experiencing difficulties, will the new venture also provide poor returns, or will it have the effect of dampening the ups and downs in the rest of the business?

Companies must take risks if they are to survive and prosper. The risk management function's primary responsibility is to understand the portfolio of risks that the company is currently taking and the risks it plans to take in the future. It must decide whether the risks are acceptable and, if they are not acceptable, what action should be taken.

Most of this book is concerned with the ways risks are managed by banks and other financial institutions, but many of the ideas and approaches we will discuss are equally applicable to nonfinancial corporations. Risk management has become progressively more important for all corporations in the last few decades. Financial institutions in particular are finding they have to increase the resources they devote to risk management. In spite of this, financial institutions still make mistakes. In March 2021, Archegos,

a hedge fund managing the assets of a rich private individual named Bill Hwang, went bankrupt as a result of highly levered risky investments. Several banks that the hedge fund had been dealing with lost huge sums. It is reported that Credit Suisse lost over \$5 billion and Nomura lost almost \$3 billion as a result of their dealings with Archegos.

Risk management is not about minimizing risks. It is about ensuring that the risks taken are manageable and that the expected returns are commensurate with the risks being taken. This chapter sets the scene. It starts by reviewing the classical arguments concerning the risk-return trade-offs faced by a well-diversified investor who is choosing a portfolio of stocks and bonds. It then considers whether the same arguments can be used by a company in choosing new projects and managing its risk exposure. The chapter concludes that there are reasons why all companies, including financial institutions, should be concerned with the total risk they face, not just with the risk from the viewpoint of a well-diversified shareholder.

1.1 Risk vs. Return for Investors

We now examine the theoretical trade-off between risk and return when money is invested. The first point to note is that the trade-off is actually between risk and *expected return*, not between risk and actual return. The term “expected return” sometimes causes confusion. In everyday language an outcome that is “expected” is considered highly likely to occur. However, statisticians define the expected value of a variable as its average (or mean) value. Expected return is therefore a weighted average of the possible returns, where the weight applied to a particular return equals the probability of that return occurring. The possible returns and their probabilities can be either estimated from historical data or assessed subjectively.

Suppose, for example, that you have \$100,000 to invest for one year. Suppose further that Treasury bills yield 5%. One alternative is to buy Treasury bills. There is then no risk and the expected return is 5%. Another alternative is to invest the \$100,000 in a stock. To simplify things, we suppose that the possible outcomes from this investment are as shown in Table 1.1. There is a 0.05 probability that the return will be +50%; there is a 0.25 probability that the return will be +30%; and so on. Expressing the returns in decimal form, the expected return per year is:

$$0.05 \times 0.50 + 0.25 \times 0.30 + 0.40 \times 0.10 + 0.25 \times (-0.10) + 0.05 \times (-0.30) = 0.10$$

This shows that, in return for taking some risk, you are able to increase your expected return per annum from the 5% offered by Treasury bills to 10%. If things work out well, your return per annum could be as high as 50%. But the worst-case outcome is a -30% return or a loss of \$30,000.

One of the first attempts to understand the trade-off between risk and expected return was by Markowitz (1952). Later, Sharpe (1964) and others carried the Markowitz

Table 1.1 One-Year Return from Investing \$100,000 in a Stock

Probability	Return
0.05	+50%
0.25	+30%
0.40	+10%
0.25	−10%
0.05	−30%

analysis a stage further by developing what is known as the capital asset pricing model. This is a relationship between expected return and what is termed *systematic risk*, which will be explained later in this chapter. In 1976, Ross developed arbitrage pricing theory, which can be viewed as an extension of the capital asset pricing model to the situation where there are several sources of systematic risk. The key insights of these researchers have had a profound effect on the way portfolio managers think about and analyze the risk-return trade-offs they face. We will now review these insights.

1.1.1 Quantifying Risk

How do you quantify the risk you are taking when you choose an investment? A convenient measure that is often used is the standard deviation of the return over one year. This is

$$\sqrt{E(R^2) - [E(R)]^2}$$

where R is the return per annum. The symbol E denotes expected value so that $E(R)$ is the expected return per annum. In Table 1.1, as we have shown, $E(R) = 0.10$. To calculate $E(R^2)$ we must weight the alternative squared returns by their probabilities:

$$\begin{aligned} E(R^2) &= 0.05 \times 0.50^2 + 0.25 \times 0.30^2 + 0.40 \times 0.10^2 + 0.25 \times (-0.10)^2 \\ &\quad + 0.05 \times (-0.30)^2 = 0.046 \end{aligned}$$

The standard deviation of the annual return is therefore $\sqrt{0.046 - 0.1^2} = 0.1897$ or 18.97%.

1.1.2 Investment Opportunities

Suppose we choose to characterize every investment opportunity by its expected return and standard deviation of return. We can plot available risky investments on a chart such as Figure 1.1, where the horizontal axis is the standard deviation of the return and the vertical axis is the expected return.

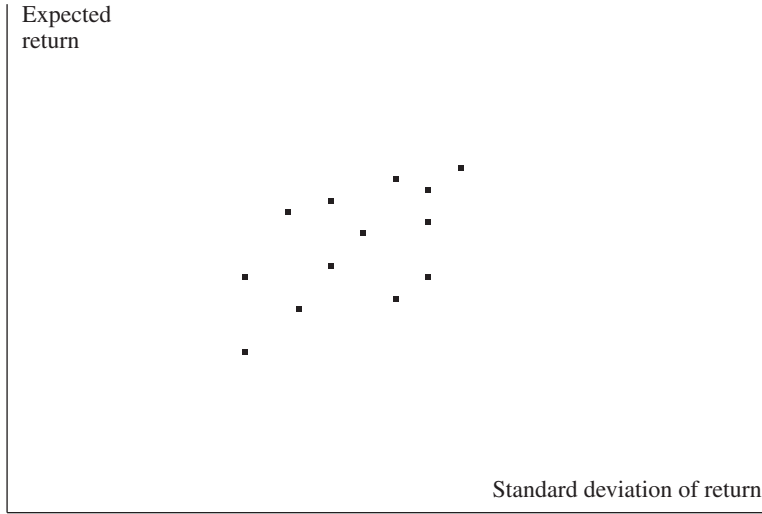


Figure 1.1 Alternative Risky Investments

Once we have identified the expected return and the standard deviation of the return for individual investments, it is natural to think about what happens when we combine investments to form a portfolio. Consider two investments with returns R_1 and R_2 . The return from putting a proportion w_1 of our money in the first investment and a proportion $w_2 = 1 - w_1$ in the second investment is

$$w_1 R_1 + w_2 R_2$$

The portfolio expected return is

$$\mu_p = w_1 \mu_1 + w_2 \mu_2 \quad (1.1)$$

where μ_1 is the expected return from the first investment and μ_2 is the expected return from the second investment. The standard deviation of the portfolio return is given by

$$\sigma_p = \sqrt{w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2\rho w_1 w_2 \sigma_1 \sigma_2} \quad (1.2)$$

where σ_1 and σ_2 are the standard deviations of R_1 and R_2 and ρ is the coefficient of correlation between the two.

Suppose that μ_1 is 10% per annum and σ_1 is 16% per annum, while μ_2 is 15% per annum and σ_2 is 24% per annum. Suppose also that the coefficient of correlation, ρ , between the returns is 0.2 or 20%. Table 1.2 shows the values of μ_p and σ_p for a number of different values of w_1 and w_2 . The calculations show that by putting part of your money in the first investment and part in the second investment a wide range of risk-return combinations can be achieved. These are plotted in Figure 1.2.

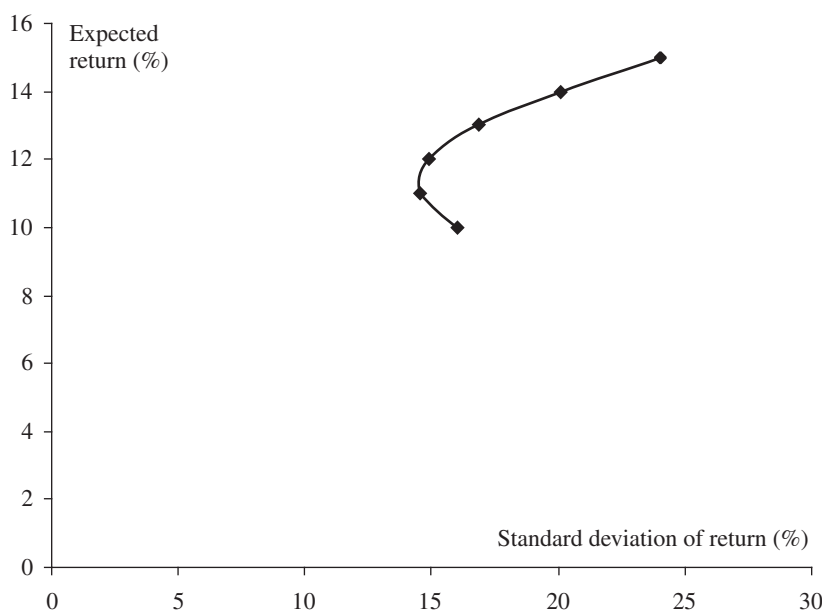


Figure 1.2 Alternative Risk-Return Combinations from Two Investments (as Calculated in Table 1.2)

Table 1.2 Expected Return, μ_p , and Standard Deviation of Return, σ_p , from a Portfolio Consisting of Two Investments

w_1	w_2	μ_p	σ_p
0.0	1.0	15%	24.00%
0.2	0.8	14%	20.09%
0.4	0.6	13%	16.89%
0.6	0.4	12%	14.87%
0.8	0.2	11%	14.54%
1.0	0.0	10%	16.00%

The expected returns from the investments are 10% and 15%; the standard deviations of the returns are 16% and 24%; and the correlation between returns is 0.2.

Most investors are risk-averse. They want to increase expected return while reducing the standard deviation of return. This means that they want to move as far as they can in a “northwest” direction in Figures 1.1 and 1.2. Figure 1.2 shows that forming a portfolio of the two investments we have been considering helps them do this. For example, by putting 60% in the first investment and 40% in the second, a portfolio with an expected return of 12% and a standard deviation of return equal to 14.87% is obtained. This is an improvement over the risk-return trade-off from putting 100% in the first investment. (The expected return is 2% higher and the standard deviation of the return is 1.13% lower.) Assuming that the investor is only concerned with the mean and standard deviation of returns and is risk averse, we can say that putting all funds in the first investment is suboptimal. It is dominated by other opportunities.

1.2 The Efficient Frontier

Let us now bring a third investment into our analysis. The third investment can be combined with any combination of the first two investments to produce new risk-return combinations. This enables us to move further in the northwest direction. We can then add a fourth investment. This can be combined with any combination of the first three investments to produce yet more investment opportunities. As we continue this process, considering every possible portfolio of the available risky investments, we obtain what is known as an *efficient frontier*. This represents the limit of how far we can move in a northwest direction and is illustrated in Figure 1.3. There is no investment that dominates a point on the efficient frontier in the sense that it has both a higher expected return and a lower standard deviation of return. The area southeast of the efficient frontier represents the set of all investments that are possible. For any point in this area that is not on the efficient frontier, there is a point on the efficient frontier that has a higher expected return and lower standard deviation of return.

In Figure 1.3 we have considered only risky investments. What does the efficient frontier of all possible investments look like? Specifically, what happens when we include the risk-free investment? Suppose that the risk-free investment yields a return of R_F . In Figure 1.4 we have denoted the risk-free investment by point F and drawn a tangent from point F to the efficient frontier of risky investments that was developed in Figure 1.3. M is the point of tangency. As we will now show, the line FJ is our new efficient frontier.

Consider what happens when we form an investment I by putting β_I of the funds we have available for investment in the risky portfolio, M , and $1 - \beta_I$ in the risk-free investment, F ($0 < \beta_I < 1$). From equation (1.1) the expected return from the investment, $E(R_I)$, is given by

$$E(R_I) = (1 - \beta_I)R_F + \beta_I E(R_M)$$

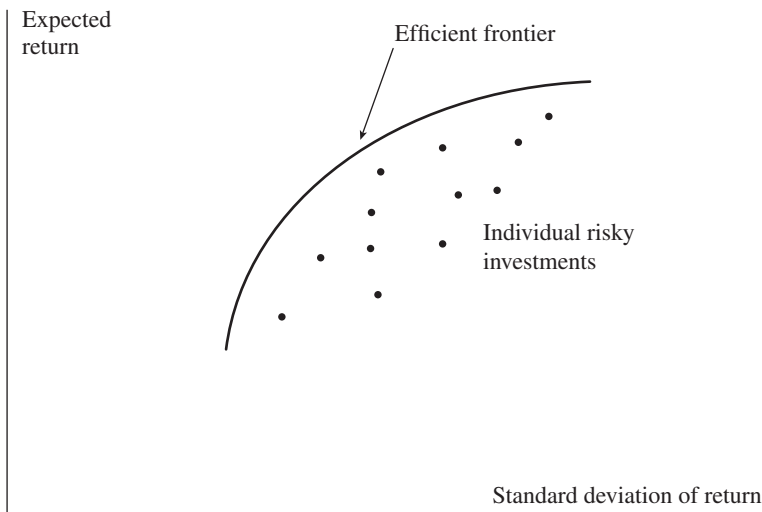


Figure 1.3 Efficient Frontier Obtainable from Risky Investments

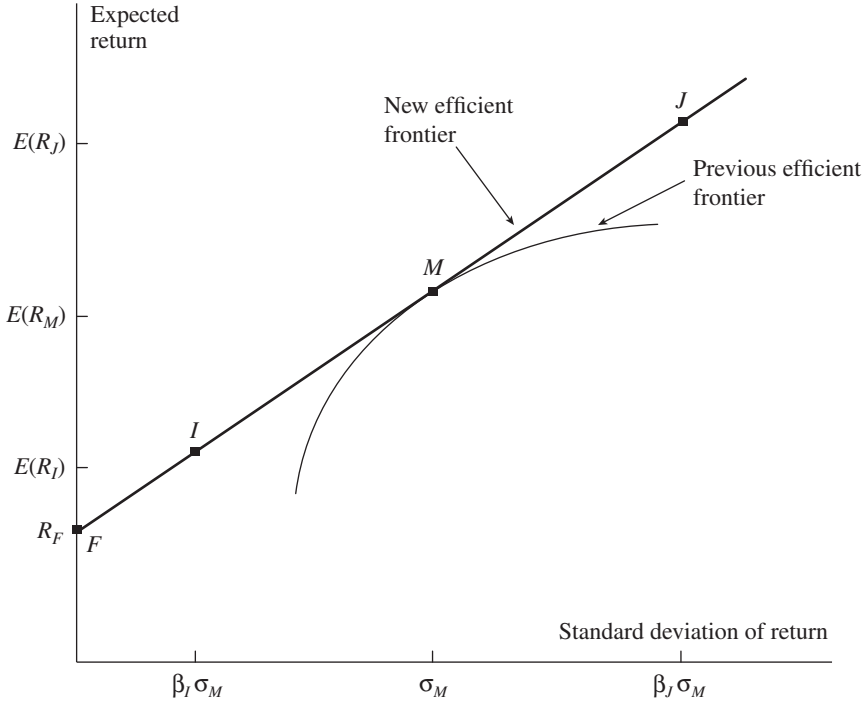


Figure 1.4 The Efficient Frontier of All Investments

Point I is achieved by investing a percentage β_I of available funds in portfolio M and the rest in a risk-free investment. Point J is achieved by borrowing $\beta_J - 1$ of available funds at the risk-free rate and investing everything in portfolio M .

and from equation (1.2), because the risk-free investment has zero standard deviation, the return R_I has standard deviation

$$\beta_I \sigma_M$$

where σ_M is the standard deviation of return for portfolio M . This risk-return combination corresponds to the point labeled I in Figure 1.4. From the perspective of both expected return and standard deviation of return, point I is β_I of the way from F to M .

All points on the line FM can be obtained by choosing a suitable combination of the investment represented by point F and the investment represented by point M . The points on this line dominate all the points on the previous efficient frontier because they give a better risk-return combination. The straight line FM is therefore part of the new efficient frontier.

If we make the simplifying assumption that we can borrow at the risk-free rate of R_F as well as invest at that rate, we can create investments that are on the continuation of FM beyond M . Suppose, for example, that we want to create the investment represented by the point J in Figure 1.4 where the distance of J from F is β_J times the distance of M from F ($\beta_J > 1$). We borrow $\beta_J - 1$ of the amount that we have available for investment at rate R_F and then invest everything (the original funds and the borrowed funds) in

the investment represented by point M . After allowing for the interest paid, the new investment has an expected return, $E(R_J)$, given by

$$E(R_J) = \beta_J E(R_M) - (\beta_J - 1)R_F = (1 - \beta_J)R_F + \beta_J E(R_M)$$

and the standard deviation of the return is

$$\beta_J \sigma_M$$

This shows that the risk and expected return combination corresponds to point J . (Note that the formulas for the expected return and standard deviation of return in terms of beta are the same whether beta is greater than or less than 1.)

The idealized argument that we have presented shows that, when the risk-free investment is considered, the efficient frontier must be a straight line. To put this another way, there should be a linear trade-off between the expected return and the standard deviation of returns, as indicated in Figure 1.4. All investors should choose the same portfolio of risky assets. This is the portfolio represented by M . They should then reflect their appetite for risk by combining this risky investment with borrowing or lending at the risk-free rate.

It is a short step from here to argue that the portfolio of risky investments represented by M must be the portfolio of all risky investments. Suppose a particular investment is not in the portfolio. No investors would hold it and its price would have to go down so that its expected return increased and it became part of portfolio M . In fact, we can go further than this. To ensure a balance between the supply and demand for each investment, the price of each risky investment must adjust so that the amount of that investment in portfolio M is proportional to the amount of that investment available in the economy. The investment represented by point M is therefore usually referred to as the *market portfolio*.

1.3 The Capital Asset Pricing Model

How do investors decide on the expected returns they require for individual investments? Based on the analysis we have presented, the market portfolio should play a key role. The expected return required on an investment should reflect the extent to which the investment contributes to the risks of the market portfolio.

A common procedure is to use historical data and regression analysis to determine a best-fit linear relationship between returns from an investment and returns from the market portfolio. This relationship has the form:

$$R = a + \beta R_M + \epsilon \tag{1.3}$$

where R is the return from the investment, R_M is the return from the market portfolio, a and β are constants, and ϵ is a random variable equal to the regression error.

Equation (1.3) shows that there are two uncertain components to the risk in the investment's return:

1. A component βR_M , which is a multiple of the return from the market portfolio.
2. A component ϵ , which is unrelated to the return from the market portfolio.

The first component is referred to as *systematic risk*. The second component is referred to as *nonsystematic risk*.

Consider first the nonsystematic risk. If we assume that the ϵ variables for different investments are independent of each other, the nonsystematic risk is almost completely diversified away in a large portfolio. An investor should not therefore be concerned about nonsystematic risk and should not require an extra return above the risk-free rate for bearing nonsystematic risk.

The systematic risk component is what should matter to a well-diversified investor. When a large well-diversified portfolio is held, the systematic risk represented by βR_M does not disappear. An investor should require an expected return to compensate for this systematic risk.

We know how investors trade off systematic risk and expected return from Figure 1.4. When $\beta = 0$ there is no systematic risk and the expected return is R_F . When $\beta = 1$, we have the same systematic risk as the market portfolio, which is represented by point M , and the expected return should be $E(R_M)$. In general

$$E(R) = R_F + \beta[E(R_M) - R_F] \quad (1.4)$$

This is the *capital asset pricing model*. The excess expected return over the risk-free rate required on the investment is β times the excess expected return on the market portfolio. This relationship is plotted in Figure 1.5. The parameter β is the *beta* of the investment.

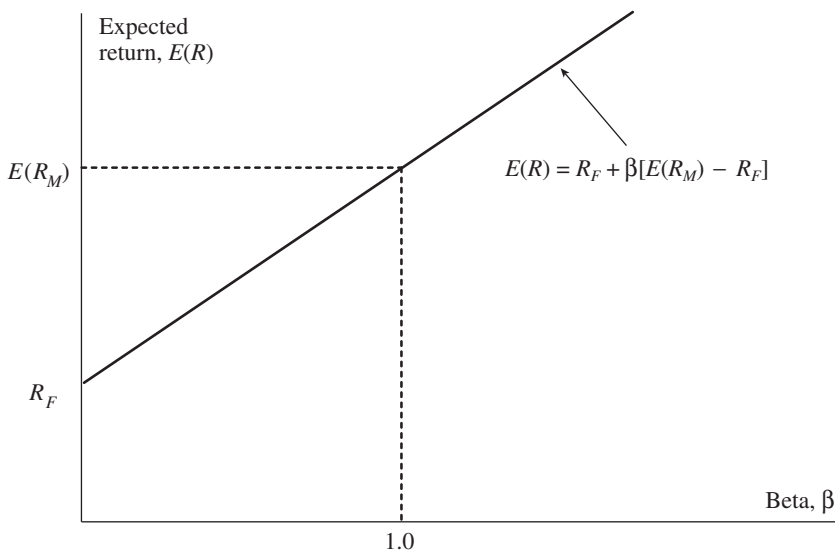


Figure 1.5 The Capital Asset Pricing Model

Example 1.1

Suppose that the risk-free rate is 5% and the return on the market portfolio is 10%. An investment with a beta of 0 should have an expected return of 5%. This is because all of the risk in the investment can be diversified away. An investment with a beta of 0.5 should have an expected return of

$$0.05 + 0.5 \times (0.1 - 0.05) = 0.075$$

or 7.5%. An investment with a beta of 1.2 should have an expected return of

$$0.05 + 1.2 \times (0.1 - 0.05) = 0.11$$

or 11%.

The parameter β is equal to $\rho\sigma/\sigma_M$, where ρ is the correlation between the return on the investment and the return on the market portfolio, σ is the standard deviation of the return on the investment, and σ_M is the standard deviation of the return on the market portfolio. Beta measures the sensitivity of the return on the investment to the return on the market portfolio. We can define the beta of any investment portfolio as in equation (1.3) by regressing its returns against the returns on the market portfolio. The capital asset pricing model in equation (1.4) should then apply with the return R defined as the return on the portfolio. In Figure 1.4 the market portfolio represented by M has a beta of 1.0 and the riskless portfolio represented by F has a beta of zero. The portfolios represented by the points I and J have betas equal to β_I and β_J , respectively.

1.3.1 Assumptions

The idealized analysis we have presented leads to the surprising conclusion that all investors want to hold the same portfolios of assets (the portfolio represented by M in Figure 1.4). This is clearly not true. Indeed, if it were true, markets would not function at all well because investors would not want to trade with each other! In practice, different investors have different views on the attractiveness of stocks and other risky investment opportunities. This is what causes them to trade with each other and it is this trading that leads to the formation of prices in markets.

The reason why the analysis leads to conclusions that do not correspond with the realities of markets is that, in presenting the arguments, we implicitly made a number of assumptions. In particular:

1. We assumed that investors care only about the expected return and the standard deviation of return of their portfolio. Another way of saying this is that investors look only at the first two moments of the return distribution. If returns are normally

distributed, it is reasonable for investors to do this. However, the returns from many assets are non-normal. They have *skewness* and *excess kurtosis*. Skewness is related to the third moment of the distribution and excess kurtosis is related to the fourth moment. In the case of positive skewness, very high returns are more likely and very low returns are less likely than the normal distribution would predict; in the case of negative skewness, very low returns are more likely and very high returns are less likely than the normal distribution would predict. Excess kurtosis leads to a distribution where both very high and very low returns are more likely than the normal distribution would predict. Most investors are concerned about the possibility of extreme negative outcomes. They are likely to want a higher expected return from investments with negative skewness or excess kurtosis.

2. We assumed that the ϵ variables for different investments in equation (1.3) are independent. Equivalently we assumed the returns from investments are correlated with each other only because of their correlation with the market portfolio. This is clearly not true. Ford and General Motors are both in the automotive sector. There is likely to be some correlation between their returns that does not arise from their correlation with the overall stock market. This means that the ϵ for Ford and the ϵ for General Motors are not likely to be independent of each other.
3. We assumed that investors focus on returns over just one period and the length of this period is the same for all investors. This is also clearly not true. Some investors such as pension funds have very long time horizons. Others such as day traders have very short time horizons.
4. We assumed that investors can borrow and lend at the same risk-free rate. This is approximately true in normal market conditions for a large financial institution that has a good credit rating. But it is not true for small investors.
5. We did not consider tax. In some jurisdictions, capital gains are taxed differently from dividends and other sources of income. Some investments get special tax treatment and not all investors are subject to the same tax rate. In practice, tax considerations have a part to play in the decisions of an investor. An investment that is appropriate for a pension fund that pays no tax might be inappropriate for a high-marginal-rate taxpayer living in New York, and vice versa.
6. Finally, and perhaps most importantly, we assumed that all investors make the same estimates of expected returns, standard deviations of returns, and correlations between returns for available investments. To put this another way, we assumed that investors have *homogeneous expectations*. This is clearly not true. Indeed, if we lived in a world of homogeneous expectations there would be no trading.

In spite of all this, the capital asset pricing model has proved to be a useful tool for portfolio managers. Estimates of the betas of stocks are readily available and the expected return on a portfolio estimated by the capital asset pricing model is a commonly used benchmark for assessing the performance of the portfolio manager, as we will now explain.

1.3.2 Alpha

When we observe a return of R_M on the market, what do we expect the return on a portfolio with a beta of β to be? The capital asset pricing model relates the expected return on a portfolio to the expected return on the market. But it can also be used to relate the expected return on a portfolio to the actual return on the market:

$$E(R_p) = R_F + \beta(R_M - R_F)$$

where R_F is the risk-free rate and R_p is the return on the portfolio.

Example 1.2

Consider a portfolio with a beta of 0.6 when the risk-free interest rate is 4%. When the return on the market is 20%, the expected return on the portfolio is

$$0.04 + 0.6 \times (0.2 - 0.04) = 0.136$$

or 13.6%. When the return on the market is 10%, the expected return from the portfolio is

$$0.04 + 0.6 \times (0.1 - 0.04) = 0.076$$

or 7.6%. When the return from the market is -10% , the expected return from the portfolio is

$$0.04 + 0.6 \times (-0.1 - 0.04) = -0.044$$

or -4.4% . The relationship between the expected return on the portfolio and the return on the market is shown in Figure 1.6.

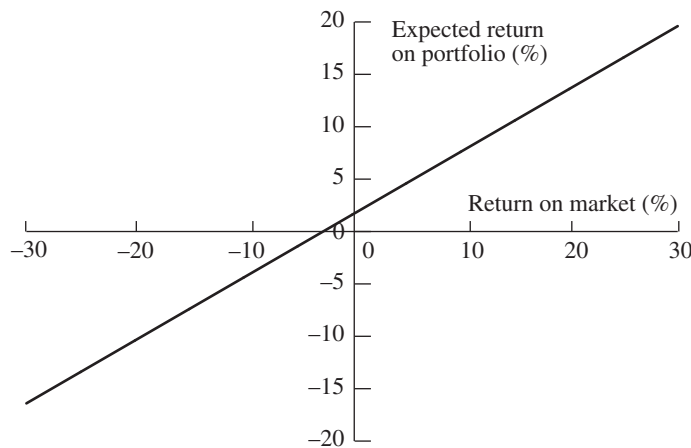


Figure 1.6 Relationship between Expected Return on Portfolio and the Actual Return on the Market When Portfolio Beta Is 0.6 and Risk-Free Rate Is 4%

Suppose that the actual return on the portfolio is greater than the expected return:

$$R_P > R_F + \beta(R_M - R_F)$$

The portfolio manager has produced a superior return for the amount of systematic risk being taken. The extra return is

$$\alpha = R_P - R_F - \beta(R_M - R_F)$$

This is commonly referred to as the *alpha* created by the portfolio manager.¹

Example 1.3

A portfolio manager has a portfolio with a beta of 0.8. The one-year risk-free rate of interest is 5%, the return on the market during the year is 7%, and the portfolio manager's return is 9%. The manager's alpha is

$$\alpha = 0.09 - 0.05 - 0.8 \times (0.07 - 0.05) = 0.024$$

or 2.4%.

Portfolio managers are continually searching for ways of producing a positive alpha. One way is by trying to pick stocks that outperform the market. Another is by *market timing*. This involves trying to anticipate movements in the market as a whole and moving funds from safe investments such as Treasury bills to the stock market when an upturn is anticipated and in the other direction when a downturn is anticipated. Chapter 4 explains other strategies used by hedge funds to try to create positive alpha.

Although the capital asset pricing model is unrealistic in some respects, the alpha and beta parameters that come out of the model are widely used to characterize investments. Beta describes the amount of systematic risk. The higher the value of beta, the greater the systematic risk being taken and the greater the extent to which returns are dependent on the performance of the market. Alpha represents the extra return made from superior portfolio management (or perhaps just good luck). An investor can make a positive alpha only at the expense of other investors who are making a negative alpha. The weighted average alpha of all investors must be zero.

1.4 Arbitrage Pricing Theory

Arbitrage pricing theory can be viewed as an extension of the capital asset pricing model. In the capital asset pricing model, an asset's return depends on just one factor. In arbitrage pricing theory, the return depends on several factors. (These factors might involve

¹It is sometimes referred to as Jensen's alpha because it was first used by Michael Jensen in evaluating mutual fund performance. See Section 4.3.

variables such as the gross national product, the domestic interest rate, and the inflation rate.) By exploring ways in which investors can form portfolios that eliminate exposure to the factors, arbitrage pricing theory shows that the expected return from an investment is linearly dependent on the factors.

The assumption that the ϵ variables for different investments are independent in equation (1.3) ensures that there is just one factor driving expected returns (and therefore one source of systematic risk) in the capital asset pricing model. This is the return on the market portfolio. In arbitrage pricing theory there are several factors affecting investment returns. Each factor is a separate source of systematic risk. Unsystematic (i.e., diversifiable) risk in arbitrage pricing theory is the risk that is unrelated to all the factors.

1.5 Risk vs. Return for Companies

We now move on to consider the trade-offs between risk and return made by a company. How should a company decide whether the expected return on a new investment project is sufficient compensation for its risks?

The ultimate owners of a company are its shareholders and a company should be managed in the best interests of its shareholders. It is therefore natural to argue that a new project undertaken by the company should be viewed as an addition to its shareholders' portfolio. The company should calculate the beta of the investment project and its expected return. If the expected return is greater than that given by the capital asset pricing model, it is a good deal for shareholders and the investment should be accepted. Otherwise it should be rejected.

The argument just given suggests that nonsystematic risks should not be considered when accept/reject decisions on new projects are taken. In practice, companies are concerned about nonsystematic as well as systematic risks. For example, most companies insure themselves against the risk of their buildings burning down—even though this risk is entirely nonsystematic and can be diversified away by their shareholders. They try to avoid taking high risks and often hedge their exposures to exchange rates, interest rates, commodity prices, and other market variables.

Earnings stability and the survival of the company are often important managerial objectives. Companies do try to ensure that their expected returns on new ventures are consistent with the risk–return trade-offs of their shareholders. But there is an overriding constraint that the total risks taken should not be allowed to get too large.

Many investors are also concerned about the overall risk of the companies they invest in. They do not like surprises and prefer to invest in companies that show solid growth and meet earnings forecasts. They like companies to manage risks carefully and limit the overall amount of risk—both systematic and nonsystematic—they are taking.

The theoretical arguments we presented in Sections 1.1 to 1.4 suggest that investors should not behave in this way. They should hold a well-diversified portfolio and encourage the companies they invest in to make high-risk investments when the combination

of expected return and systematic risk is favorable. Some of the companies in a shareholder's portfolio will go bankrupt, but others will do very well. The result should be an overall return to the shareholder that is satisfactory.

Are investors behaving suboptimally? Would their interests be better served if companies took more nonsystematic risks? There is an important argument to suggest that this is not necessarily the case. This argument is usually referred to as the “bankruptcy costs” argument. It is often used to explain why a company should restrict the amount of debt it takes on, but it can be extended to apply to a wider range of risk management decisions than this.

1.5.1 Bankruptcy Costs

In a perfect world, bankruptcy would be a fast affair where the company's assets (tangible and intangible) are sold at their fair market value and the proceeds are distributed to the company's creditors using well-defined rules. If we lived in such a perfect world, the bankruptcy process itself would not destroy value for stakeholders. Unfortunately, the real world is far from perfect. By the time a company reaches the point of bankruptcy, it is likely that its assets have lost some value. The bankruptcy process itself invariably reduces the value of its assets further. This further reduction in value is referred to as *bankruptcy costs*.

What is the nature of bankruptcy costs? Once a bankruptcy has happened, customers and suppliers become less inclined to do business with the company; assets sometimes have to be sold quickly at prices well below those that would be realized in an orderly sale; the value of important intangible assets, such as the company's brand name and its reputation in the market, are often destroyed; the company is no longer run in the best interests of shareholders; large fees are often paid to accountants and lawyers; and so on. The story in Business Snapshot 1.1 is representative of what often happens in practice. It illustrates how, when a high-risk decision works out badly, there can be disastrous bankruptcy costs.

The largest bankruptcy in U.S. history was that of Lehman Brothers on September 15, 2008. The bankruptcy proceedings continued for many years. Denison, Fleming and Sarkar at the Federal Reserve Bank of New York estimated in 2019 that the direct and indirect costs of the bankruptcy totaled between \$46 billion and \$63 billion, or 15% to 21% of Lehman's pre-bankruptcy consolidated assets.²

We mentioned earlier that corporate survival is an important managerial objective and that shareholders like companies to avoid excessive risks. We now understand one reason why this is so. Bankruptcy laws vary widely from country to country, but they all have the effect of destroying value as lenders and other creditors vie with each other to get

²See: <https://libertystreeteconomics.newyorkfed.org/2019/01/how-much-value-was-destroyed-by-the-lehman-bankruptcy/>.

paid. If a company chooses projects with very high risks (but sufficiently high expected returns to be above the efficient frontier in Figure 1.4), the probability of bankruptcy will be quite high. When expected bankruptcy costs are taken into account, projects that have a high total (systematic plus nonsystematic) risk are liable to be rejected as unacceptable. This explains why investors like companies to limit the overall amount of risk they take and reward companies that manage risks so that they meet earnings forecasts.

BUSINESS SNAPSHOT 1.1

The Hidden Costs of Bankruptcy: A Hypothetical Scenario

Several years ago, a company had a market capitalization of \$2 billion and \$500 million of debt. The CEO decided to acquire a company in a related industry for \$1 billion in cash. The cash was raised using a mixture of bank debt and bond issues. The price paid for the company was justified on the basis of potential synergies, but key threats to the profitability of the company were overlooked.

Many of the anticipated synergies were not realized. Furthermore, the company that was acquired was not profitable and proved to be a cash drain on the parent company. After three years, the CEO resigned. The new CEO sold the acquisition for \$100 million (10% of the price paid) and announced that the company would focus on its original core business. However, by then the company was highly leveraged. A temporary economic downturn made it impossible for the company to service its debt and it declared bankruptcy.

The offices of the company were soon filled with accountants and lawyers representing the interests of the various parties (banks, different categories of bondholders, equity holders, the company, and the board of directors). These people directly or indirectly billed the company about \$10 million per month in fees. The company lost sales that it would normally have made because nobody wants to do business with a bankrupt company. Intangible assets, such as the relationships built up over many years with customers, were lost. Key senior executives, who had many years of valuable experience, left.

After two years and three reorganization attempts, an agreement was reached among the various parties, and a new company with a market capitalization of \$700,000 was incorporated to continue the remaining profitable parts of the business. The shares in the new company were entirely owned by the banks and the bondholders. The shareholders got nothing.

When a major new investment is being contemplated, it is important to consider how well it fits in with other risks taken by the company. Relatively small investments can often have the effect of reducing the overall risks taken because of their diversification benefits. However, a large investment can dramatically increase these risks. Many spectacular corporate failures (such as the fictitious one in Business Snapshot 1.1) can be traced to CEOs who made large acquisitions (often highly leveraged) that did not work out.

1.5.2 Financial Institutions

One can argue about how important bankruptcy costs are for the decision making of a non-financial company, but there can be no question that it is crucially important for a financial institution such as a bank to keep its probability of bankruptcy very low. Large banks rely on wholesale deposits and instruments such as commercial paper for their funding. Confidence is the key to their survival. If the risk of default is perceived by the market to be other than very low, there will be a lack of confidence and sources of funding will dry up. The bank will then be forced into liquidation—even if it is solvent in the sense of having positive equity. Lehman Brothers was the largest bankruptcy in U.S. history. Northern Rock was a large failure of a financial institution in the United Kingdom. In both cases, the failure was because there was a lack of confidence and traditional sources of funding dried up.

1.5.3 Regulation

Even if, in spite of the arguments we have just given, the managers of a bank wanted to take huge risks, they would not be allowed to do so. Unlike other companies, many financial institutions are heavily regulated. Governments throughout the world want a stable financial sector. It is important that companies and private individuals have confidence in banks and insurance companies when they transact business. The regulations are designed to ensure that the probability of a large bank or an insurance company experiencing severe financial difficulties is low. The bailouts of financial institutions in 2008 during the Global Financial Crisis illustrate the reluctance of governments to let large financial institutions fail. Regulated financial institutions are forced to consider total risks (systematic plus nonsystematic).

Bankruptcy generally arises from losses being incurred. Regulators try to ensure that the capital held by a bank is sufficient to provide a cushion to absorb the losses with a high probability. Suppose, for example, that there is considered to be only a 0.1% probability that a financial institution will experience a loss of \$2 billion or more in a year. Regulators might require the bank to hold equity capital equal to \$2 billion. This would ensure that there is a 99.9% probability that the equity capital is sufficient to absorb the losses. The models used by regulators are discussed in more detail in later chapters.

The key point here is that regulators are concerned with total risks, not just systematic risks. Their goal is to make bankruptcy a highly unlikely event.

1.6 Risk Management by Financial Institutions

There are two broad risk management strategies open to a financial institution (or any other organization). One approach is to identify risks one by one and handle each one separately. This is sometimes referred to as *risk decomposition*. The other is to reduce risks by being well diversified. This is sometimes referred to as *risk aggregation*. Both approaches are typically used by financial institutions.

Consider, for example, the market risks incurred by the trading room of a bank. These risks depend on the future movements in a multitude of market variables (exchange rates, interest rates, stock prices, and so on). To implement the risk decomposition approach, the trading room is organized so that a trader is responsible for trades related to just one market variable (or perhaps a small group of market variables). For example, there could be one trader who is responsible for all trades involving the dollar-yen exchange rate. At the end of each day, the trader is required to ensure that certain risk measures are kept within limits specified by the bank. If the end of the day is approached and it looks as though one or more of the risk measures will be outside the specified limits, the trader must either get special permission to maintain the position or execute new hedging trades so that the limits are adhered to. (The risk measures and the way they are used are discussed in later chapters.)

The risk managers, working in what is termed the middle office of a bank, implement the risk aggregation approach for the market risks being taken. This involves calculating at the end of each day the total risk faced by the bank from movements in all market variables. Hopefully, the bank is well diversified so that its overall exposure to market movements is fairly small. If risks are unacceptably high, then the reasons must be determined and corrective action taken.

Risk aggregation is a key tool for insurance companies. Consider automobile insurance. The insurance company's payout on one particular automobile insurance policy is quite uncertain. However, the payout on 100,000 similar insurance policies can be predicted with reasonable accuracy.

Credit risks are also traditionally managed using risk aggregation. It is important for a financial institution to be well diversified. If, for example, a bank lends 40% of its available funds to a single borrower, it is not well diversified and likely to be subject to unacceptable risks. If the borrower runs into financial difficulties and is unable to make interest and principal payments, the bank could become insolvent.

If the bank adopts a more diversified strategy of lending 0.01% of its available funds to each of 10,000 different borrowers, it is in a much safer position. Suppose that the probability of any one borrower defaulting is 1%. We can expect that close to 100 borrowers will default in the year and the losses on these borrowers will be more than offset by the profits earned on the 99% of loans that perform well. To maximize the benefits of diversification, borrowers should be in different geographical regions and different industries. A large international bank with different types of borrowers all over the world is likely to be much better diversified than a small bank in Texas that lends entirely to oil companies.

But, however well diversified a bank is, it is still exposed to systematic risk, which creates variations in the probability of default for all borrowers from year to year. Suppose that the probability of default for borrowers in an average year is 1%. When the economy is doing well, the probability of default is less than this and when there is an economic downturn it is liable to be considerably more than this. Models for capturing this exposure are discussed in later chapters.

Since the late 1990s, we have seen the emergence of an active market for credit derivatives. Credit derivatives allow banks to handle credit risks one by one (risk decomposition) rather than relying solely on risk diversification. They also allow banks to buy protection against the overall level of defaults in the economy. However, for every buyer of credit protection there must be a seller. Many sellers of credit protection, whether on individual companies or on portfolios of companies, took huge losses during the Global Financial Crisis that started in 2007. (This crisis is discussed further in Chapter 7.)

1.7 Credit Ratings

Credit rating agencies provide information that is widely used by financial market participants for the management of credit risks. A credit rating is a measure of the credit quality of a debt instrument such as a bond. However, the rating of a corporate or sovereign bond is often assumed to be an attribute of the bond issuer rather than of the bond itself. Thus, if the bonds issued by a company have a rating of AAA, the company is often referred to as having a rating of AAA.

The three major credit rating agencies are Moody's, S&P, and Fitch. The best rating assigned by Moody's is Aaa. Bonds with this rating are considered to have almost no chance of defaulting. The next best rating is Aa. Following that come A, Baa, Ba, B, Caa, Ca, and C. The S&P ratings corresponding to Moody's Aaa, Aa, A, Baa, Ba, B, Caa, Ca, and C are AAA, AA, A, BBB, BB, B, CCC, CC, and C, respectively. To create finer rating measures Moody's divides the Aa rating category into Aa1, Aa2, and Aa3; it divides A into A1, A2, and A3; and so on. Similarly S&P divides its AA rating category into AA+, AA, and AA–; it divides its A rating category into A+, A, and A–; and so on. Moody's Aaa rating category and S&P's AAA rating are not subdivided, nor usually are the two lowest rating categories. Fitch's rating categories are similar to those of S&P.

There is usually assumed to be an equivalence between the meanings of the ratings assigned by the different agencies. For example, a BBB+ rating from S&P is considered equivalent to a Baa1 rating from Moody's. Instruments with ratings of BBB– (Baa3) or above are considered to be *investment grade*. Those with ratings below BBB– (Baa3) are termed *noninvestment grade* or *speculative grade* or *junk bonds*.

We will learn a lot more about credit ratings in later chapters of this book. For example, Chapter 7 discusses the role of ratings in the Global Financial Crisis. Chapters 25 and 26 provide information on how ratings are used in regulation. Chapter 17 provides statistics on the default rates of companies with different credit ratings. Chapter 19 examines the extent to which the credit ratings of companies change through time.

Summary

An important general principle in finance is that there is a trade-off between risk and return. Higher expected returns can usually be achieved only by taking higher risks. In theory, shareholders should not be concerned with risks they can diversify away. The expected return they require should reflect only the amount of systematic (i.e., non-diversifiable) risk they are bearing.

Companies, although sensitive to the risk-return trade-offs of their shareholders, are concerned about total risks when they do risk management. They do not ignore the unsystematic risk that their shareholders can diversify away. One valid reason for this is the existence of bankruptcy costs, which are the costs to shareholders resulting from the bankruptcy process itself.

For financial institutions such as banks and insurance companies there is another important reason: regulation. The regulators of financial institutions are primarily concerned with minimizing the probability that the institutions they regulate will fail. The probability of failure depends on the total risks being taken, not just the risks that cannot be diversified away by shareholders. As we will see later in this book, regulators aim to ensure that financial institutions keep enough capital for the total risks they are taking.

Two general approaches to risk management are risk decomposition and risk aggregation. Risk decomposition involves managing risks one by one. Risk aggregation involves relying on the power of diversification to reduce risks. Banks use both approaches to manage market risks. Credit risks have traditionally been managed using risk aggregation, but with the advent of credit derivatives the risk decomposition approach can be used.

Further Reading

Markowitz, H. "Portfolio Selection." *Journal of Finance* 7, no. 1 (March 1952): 77–91.

Ross, S. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* 13, no. 3 (December 1976): 341–360.

Sharpe, W. "Capital Asset Prices: A Theory of Market Equilibrium." *Journal of Finance* 19, no. 3 (September 1964): 425–442.

Smith, C. W., and R. M. Stulz. "The Determinants of a Firm's Hedging Policy." *Journal of Financial and Quantitative Analysis* 20 (1985): 391–406.

Stulz, R. M. *Risk Management and Derivatives*. Mason, OH: South-Western, 2003.

Practice Questions and Problems (Answers at End of Book)

- 1.1 An investment has probabilities 0.1, 0.2, 0.35, 0.25, and 0.1 of giving returns equal to 40%, 30%, 15%, –5%, and –15%. What are the expected returns and the standard deviations of returns?

- 1.2 Suppose that there are two investments with the same probability distribution of returns as in Problem 1.1. The correlation between the returns is 0.15. What is the expected return and standard deviation of return from a portfolio where money is divided equally between the investments?
- 1.3 For the two investments considered in Figure 1.2 and Table 1.2, what are the alternative risk–return combinations if the correlation is (a) 0.3, (b) 1.0, and (c) -1.0 ?
- 1.4 What is the difference between systematic and nonsystematic risk? Which is more important to an equity investor? Which can lead to the bankruptcy of a corporation?
- 1.5 Outline the arguments leading to the conclusion that all investors should choose the same portfolio of risky investments. What are the key assumptions?
- 1.6 The expected return on the market portfolio is 12% and the risk-free rate is 6%. What is the expected return on an investment with a beta of (a) 0.2, (b) 0.5, and (c) 1.4?
- 1.7 “Arbitrage pricing theory is an extension of the capital asset pricing model.” Explain this statement.
- 1.8 “The capital structure decision of a company is a trade-off between bankruptcy costs and the tax advantages of debt.” Explain this statement.
- 1.9 What is meant by risk aggregation and risk decomposition? Which requires an in-depth understanding of individual risks? Which requires a detailed knowledge of the correlations between risks?
- 1.10 A bank’s operational risk includes the risk of very large losses because of employee fraud, natural disasters, litigation, and so on. Do you think operational risk is best handled by risk decomposition or risk aggregation? (Operational risk will be discussed in Chapter 20.)
- 1.11 A bank’s profit next year will be normally distributed with a mean of 0.6% of assets and a standard deviation of 1.5% of assets. The bank’s equity is 4% of assets. What is the probability that the bank will have a positive equity at the end of the year? Ignore taxes.
- 1.12 Why do you think that banks are regulated to ensure that they do not take too much risk but most other companies (for example, those in manufacturing and retailing) are not?
- 1.13 List the bankruptcy costs incurred by the company in Business Snapshot 1.1.
- 1.14 The return from the market last year was 10% and the risk-free rate was 5%. A hedge fund manager with a beta of 0.6 has an alpha of 4%. What return did the hedge fund manager earn?

Further Questions

- 1.15 Suppose that one investment has a mean return of 8% and a standard deviation of return of 14%. Another investment has a mean return of 12% and a standard

deviation of return of 20%. The correlation between the returns is 0.3. Produce a chart similar to Figure 1.2 showing alternative risk-return combinations from the two investments.

- 1.16 The expected return on the market is 12% and the risk-free rate is 7%. The standard deviation of the return on the market is 15%. One investor creates a portfolio on the efficient frontier with an expected return of 10%. Another creates a portfolio on the efficient frontier with an expected return of 20%. What is the standard deviation of the return on each of the two portfolios?
- 1.17 A bank estimates that its profit next year is normally distributed with a mean of 0.8% of assets and the standard deviation of 2% of assets. How much equity (as a percentage of assets) does the company need to be (a) 99% sure that it will have a positive equity at the end of the year and (b) 99.9% sure that it will have positive equity at the end of the year? Ignore taxes.
- 1.18 A portfolio manager has maintained an actively managed portfolio with a beta of 0.2. During the last year, the risk-free rate was 5% and major equity indices performed very badly, providing returns of about -30% . The portfolio manager produced a return of -10% and claims that in the circumstances it was good. Discuss this claim.

Part One

FINANCIAL INSTITUTIONS

Chapter 2

Banks

The word “bank” originates from the Italian word *banco*. This is a desk or bench, covered by a green tablecloth, that was used several hundred years ago by Florentine bankers. The traditional role of banks has been to take deposits and make loans. The interest charged on the loans is greater than the interest paid on deposits. The difference between the two has to cover administrative costs and loan losses (i.e., losses when borrowers fail to make the agreed payments of interest and principal), while providing a satisfactory return on equity.

Today, most large banks engage in both commercial and investment banking. Commercial banking involves, among other things, the deposit-taking and lending activities we have just mentioned. Investment banking is concerned with assisting companies in raising debt and equity, and providing advice on mergers and acquisitions, major corporate restructurings, and other corporate finance decisions. Large banks are also often involved in securities trading (e.g., by providing brokerage services).

Commercial banking can be classified as *retail banking* or *wholesale banking*. Retail banking, as its name implies, involves taking relatively small deposits from private individuals or small businesses and making relatively small loans to them. Wholesale banking involves the provision of banking services to medium and large corporate clients, fund managers, and other financial institutions. Both loans and deposits are much larger in wholesale banking than in retail banking.

Typically the spread between the cost of funds and the lending rate is smaller for wholesale banking than for retail banking. However, this tends to be offset by lower costs. (When a certain dollar amount of wholesale lending is compared to the same dollar amount of retail lending, the expected loan losses and administrative costs are

usually much less.) Banks that fund their lending by borrowing in financial markets are referred to as *money center banks*.

This chapter will review how commercial and investment banking have evolved in the United States over the last hundred years. It will take a first look at the way the banks are regulated, the nature of the risks facing the banks, and the key role of capital in providing a cushion against losses.

2.1 Commercial Banking

Commercial banking in virtually all countries has been subject to a great deal of regulation. This is because most national governments consider it important that individuals and companies have confidence in the banking system. Among the issues addressed by regulation is the capital that banks must keep, the activities they are allowed to engage in, deposit insurance, and the extent to which mergers and foreign ownership are allowed. The nature of bank regulation during the twentieth century has influenced the structure of commercial banking in different countries. To illustrate this, we consider the case of the United States.

The United States is unusual in that it has a large number of banks (4,914 in 2021). This leads to a relatively complicated payment system compared with that of other countries with fewer banks. There are a few large money center banks such as Citigroup and JPMorgan Chase. There are several hundred regional banks that engage in a mixture of wholesale and retail banking, and several thousand community banks that specialize in retail banking.

Table 2.1 shows information produced by the Federal Deposit Insurance Corporation at the ends of the first quarter of 1984 and the first quarter of 2022. It can be seen that the number of banks and savings institutions declined by about 73% during this period. Small community banks have been disappearing and large banks account for an increasing share of the assets.

The structure of banking in the United States is largely a result of regulatory restrictions on interstate banking. At the beginning of the twentieth century, most U.S. banks had a single branch from which they served customers. During the early part of the twentieth century, many of these banks expanded by opening more branches in order

Table 2.1 Commercial Banking and Savings Institutions in United States, 1984 and 2022

Size (Assets)	1984		2022	
	Number of Institutions	Percent of Assets	Number of Institutions	Percent of Assets
Under \$100 million	14,034	14.30	784	0.20
\$100 million to \$1 billion	3,398	25.93	3,027	4.63
\$1 billion to \$10 billion	425	31.97	825	9.19
\$10 billion to \$250 billion	28	27.80	147	29.55
Over \$250 billion	0	0.00	13	56.43
Total	17,885	100.00	4,796	100.00

SOURCE: FDIC Quarterly Banking Profile, www.fdic.gov.

to serve their customers better. This ran into opposition from two quarters. First, small banks that still had only a single branch were concerned that they would lose market share. Second, large money center banks were concerned that the multibranch banks would be able to offer check-clearing and other payment services and erode the profits that they themselves made from offering these services. As a result, there was pressure to control the extent to which community banks could expand. Several states passed laws restricting the ability of banks to open more than one branch within a state.

The McFadden Act was passed in 1927 and amended in 1933. This act had the effect of restricting all banks from opening branches in more than one state. This restriction applied to nationally chartered as well as to state-chartered banks. One way of getting around the McFadden Act was to establish a *multibank holding company*. This is a company that acquires more than one bank as a subsidiary. By 1956, there were 47 multibank holding companies. This led to the Douglas Amendment to the Bank Holding Company Act. This did not allow a multibank holding company to acquire a bank in a state that prohibited out-of-state acquisitions. However, acquisitions prior to 1956 were grandfathered (that is, multibank holding companies did not have to dispose of acquisitions made prior to 1956).

Banks are creative in finding ways around regulations—particularly when it is profitable for them to do so. After 1956, one approach was to form a one-bank holding company. This is a holding company with just one bank as a subsidiary and a number of nonbank subsidiaries in different states from the bank. The nonbank subsidiaries offered financial services such as consumer finance, data processing, and leasing and were able to create a presence for the bank in other states.

The 1970 Bank Holding Companies Act restricted the activities of one-bank holding companies. They were only allowed to engage in activities that were closely related to banking, and acquisitions by them were subject to approval by the Federal Reserve. They had to divest themselves of acquisitions that did not conform to the act.

After 1970, the interstate banking restrictions started to disappear. Individual states passed laws allowing banks from other states to enter and acquire local banks. (Maine was the first to do so, in 1978.) Some states allowed free entry of other banks. Some allowed banks from other states to enter only if there were reciprocal agreements. (This means that state A allowed banks from state B to enter only if state B allowed banks from state A to do so.) In some cases, groups of states developed regional banking pacts that allowed interstate banking.

In 1994, the U.S. Congress passed the Riegel-Neal Interstate Banking and Branching Efficiency Act. This Act led to full interstate banking becoming a reality. It permitted bank holding companies to acquire branches in other states. It invalidated state laws that allowed interstate banking on a reciprocal or regional basis. Starting in 1997, bank holding companies were allowed to convert out-of-state subsidiary banks into branches of a single bank. Many people argued that this type of consolidation was necessary to enable U.S. banks to be large enough to compete internationally. The Riegel-Neal Act prepared the way for a wave of consolidation in the U.S. banking system (for example,

Table 2.2 Summary Balance Sheet for DLC at End of 2023 (\$ millions)

Assets		Liabilities and Net Worth	
Cash	5	Deposits	90
Marketable Securities	10	Subordinated Long-Term Debt	5
Loans	80	Equity Capital	5
Fixed Assets	5		
Total	100	Total	100

the acquisition by JPMorgan of banks formerly named Chemical, Chase, Bear Stearns, and Washington Mutual).

As a result of the Global Financial Crisis that started in 2007 and led to a number of bank failures, the Dodd–Frank Wall Street Reform and Consumer Protection Act was signed into law by President Barack Obama on July 21, 2010. This is discussed further in Chapter 26.

2.2 The Capital Requirements of a Small Commercial Bank

To illustrate the role of capital in banking, we consider a hypothetical small community bank named Deposits and Loans Corporation (DLC). DLC is primarily engaged in the traditional banking activities of taking deposits and making loans. A summary balance sheet for DLC at the end of 2023 is shown in Table 2.2 and a summary income statement for 2023 is shown in Table 2.3.

Table 2.2 shows that the bank has \$100 million in assets. Most of the assets (80% of the total) are loans made by the bank to private individuals and small corporations. Cash and marketable securities account for a further 15% of the assets. The remaining 5% of the assets are fixed assets (i.e., buildings, equipment, etc.). A total of 90% of the funding for the assets comes from deposits of one sort or another from the bank’s customers. A further 5% is financed by subordinated long-term debt. (These are bonds issued by the bank to investors that rank below deposits in the event of a liquidation.) The remaining 5% is financed by the bank’s shareholders in the form of equity capital. The equity capital consists of the original cash investment of the shareholders and earnings retained in the bank.

Consider next the income statement for 2023 shown in Table 2.3. The first item on the income statement is net interest income. This is the excess of the interest earned over the interest paid and is 3% of the total assets in our example. It is important for the bank to be managed so that net interest income remains roughly constant regardless of

Table 2.3 Summary Income Statement for DLC in 2023 (\$ millions)

Net Interest Income	3.00
Loan Losses	(0.80)
Non-Interest Income	0.90
Non-Interest Expense	(2.50)
Pre-Tax Operating Income	0.60

movements in interest rates of different maturities. We will discuss this in more detail in Chapter 14.

The next item is loan losses. This is 0.8% of total assets for the year in question. Clearly it is very important for management to quantify credit risks and manage them carefully. But however carefully a bank assesses the financial health of its clients before making a loan, it is inevitable that some borrowers will default. This is what leads to loan losses. The percentage of loans that default will tend to fluctuate from year to year with economic conditions. It is likely that in some years default rates will be quite low, while in others they will be quite high.

The next item, non-interest income, consists of income from all the activities of the bank other than lending money. This includes fees for the services the bank provides for its clients. In the case of DLC non-interest income is 0.9% of assets.

The final item is non-interest expense and is 2.5% of assets in our example. This consists of all expenses other than interest paid. It includes salaries, technology-related costs, and other overheads. As in the case of all large businesses, these have a tendency to increase over time unless they are managed carefully. Banks must try to avoid large losses from litigation, business disruption, employee fraud, and so on. The risk associated with these types of losses is known as *operational risk* and will be discussed in Chapter 20.

2.2.1 Capital Adequacy

One measure of the performance of a bank is return on equity (ROE). Tables 2.2 and 2.3 show that DLC's before-tax ROE is 0.6/5 or 12%. If this is considered unsatisfactory, one way DLC might consider improving its ROE is by buying back its shares and replacing them with deposits so that equity financing is lower and ROE is higher. For example, if it moved to the extreme balance sheet in Table 2.4 where equity is reduced to 1% of assets and deposits are increased to 94% of assets, its before-tax ROE would jump to 60%.

How much equity capital does DLC need? This question can be answered by hypothesizing an extremely adverse scenario and considering whether the bank would survive. Suppose that there is a severe recession and, as a result, the bank's loan losses rise by 3.2% of assets to 4% next year. (We assume that other items on the income statement in Table 2.3 are unaffected.) The result will be a pre-tax net operating loss of 2.6% of assets

Table 2.4 Alternative Balance Sheet for DLC at End of 2023 with Equity Only 1% of Assets (\$ millions)

Assets		Liabilities and Net Worth	
Cash	5	Deposits	94
Marketable Securities	10	Subordinated Long-Term Debt	5
Loans	80	Equity Capital	1
Fixed Assets	5		
Total	100	Total	100

($0.6 - 3.2 = -2.6$). Assuming a tax rate of 30%, this would result in an after-tax loss of about 1.8% of assets.¹

In Table 2.2, equity capital is 5% of assets, so an after-tax loss equal to 1.8% of assets, although not at all welcome, can be absorbed. It would result in a reduction of the equity capital to 3.2% of assets. Even a second bad year similar to the first would not totally wipe out the equity.

If DLC has moved to the more aggressive capital structure shown in Table 2.4, it is far less likely to survive. One year where the loan losses are 4% of assets would totally wipe out equity capital and the bank would find itself in serious financial difficulties. It would no doubt try to raise additional equity capital, but it is likely to find this difficult when in such a weak financial position. It is possible that there would be a run on the bank (where all depositors decide to withdraw funds at the same time) and the bank would be forced into liquidation. If all assets could be liquidated for book value (a big assumption), the long-term debt-holders would likely receive about \$4.2 million rather than \$5 million (they would in effect absorb the negative equity) and the depositors would be repaid in full.

Clearly, it is inadequate for a bank to have only 1% of assets funded by equity capital. Maintaining equity capital equal to 5% of assets as in Table 2.2 is more reasonable. Note that equity and subordinated long-term debt are both sources of capital. Equity provides the best protection against adverse events. (In our example, when the bank has \$5 million of equity capital rather than \$1 million, it stays solvent and is unlikely to be liquidated.) Subordinated long-term debt-holders rank below depositors in the event of default, but subordinated debt does not provide as good a cushion for the bank as equity because it does not prevent the bank's insolvency.

As we shall see in Chapters 25 to 27, bank regulators have tried to ensure that the capital a bank keeps is sufficient to cover the risks it takes. The risks include market risks, credit risks, and operational risks.

2.3 Deposit Insurance

To maintain confidence in banks, government regulators in many countries have introduced guaranty programs. These typically insure depositors against losses up to a certain level.

The United States with its large number of small banks is particularly prone to bank failures. After the stock market crash of 1929 the United States experienced a major recession and about 10,000 banks failed between 1930 and 1933. Runs on banks and panics were common. In 1933, the United States government created the Federal Deposit Insurance Corporation (FDIC) to provide protection for depositors. Originally, the maximum level of protection provided was \$2,500. This has been increased several times and became \$250,000 per depositor per bank in October 2008. Banks pay an insurance

¹This assumes that tax losses can be carried back to offset previous profits.

premium that is a percentage of their domestic deposits. The premium as a percentage of deposits depends on the bank's capital and how safe it is considered to be by regulators.

Up to 1980, the system worked well. There were no runs on banks and few bank failures. However, between 1980 and 1990, bank failures in the United States accelerated, with the total number of failures during this decade being over 1,000 (larger than for the whole 1933 to 1979 period). There were several reasons for this. One was the way in which banks managed interest rate risk and we will talk about that in Chapter 14. Another reason was the reduction in oil and other commodity prices, which led to many loans to oil, gas, and agricultural companies not being repaid.

A further reason for the bank failures was that the existence of deposit insurance allowed banks to follow risky strategies that would not otherwise be feasible. For example, they could increase their deposit base by offering high rates of interest to depositors and use the funds to make risky loans. Without deposit insurance, a bank could not follow this strategy because its depositors would see what the bank was doing, decide that the bank was too risky, and withdraw their funds. With deposit insurance, it can follow the strategy because depositors know that, if the worst happens, they are protected under FDIC. This is an example of what is known as *moral hazard*. We will talk about moral hazard further in Chapter 3. It can be defined as the possibility that the existence of insurance changes the behavior of the insured party. The introduction of risk-based deposit insurance premiums has reduced moral hazard to some extent.

During the 1980s, the funds of FDIC became seriously depleted and it had to borrow \$30 billion from the U.S. Treasury. In December 1991, Congress passed the FDIC Improvement Act to prevent any possibility of the fund becoming insolvent in the future. Between 1991 and 2006, bank failures in the United States were relatively rare and by 2006 the fund had reserves of about \$50 billion. FDIC funds were again depleted by the banks that failed as a result of the Global Financial Crisis that started in 2007. However, by 2022 the reserves had been built up again.

2.4 Investment Banking

The main activity of investment banking is raising debt and equity financing for corporations or governments. This involves originating the securities, underwriting them, and then placing them with investors. In a typical arrangement a corporation approaches an investment bank indicating that it wants to raise a certain amount of financing in the form of debt, equity, or hybrid instruments such as convertible bonds. The securities are originated complete with legal documentation itemizing the rights of the security holder. A prospectus is created outlining the company's past performance and future prospects. The risks faced by the company from such things as major lawsuits are included. There is a "road show" in which the investment bank and senior management from the company attempt to market the securities to large fund managers. A price for the securities is agreed between the bank and the corporation. The bank then sells the securities in the market.

There are a number of different types of arrangements between the investment bank and the corporation. Sometimes the financing takes the form of a *private placement* in which the securities are sold to a small number of large institutional investors, such as life insurance companies or pension funds, and the investment bank receives a fee. On other occasions it takes the form of a *public offering*, where securities are offered to the general public. A public offering may be on a *best efforts* or *firm commitment* basis. In the case of a best efforts public offering, the investment bank does as well as it can to place the securities with investors and is paid a fee that depends, to some extent, on its success. In the case of a firm commitment public offering, the investment bank agrees to buy the securities from the issuer at a particular price and then attempts to sell them in the market for a slightly higher price. It makes a profit equal to the difference between the price at which it sells the securities and the price it pays the issuer. If for any reason it is unable to sell the securities, it ends up owning them itself. The difference between the two arrangements is illustrated in Example 2.1.

Example 2.1

A bank has agreed to underwrite an issue of 50 million shares by ABC Corporation. In negotiations between the bank and the corporation the target price to be received by the corporation has been set at \$30 per share. This means that the corporation is expecting to raise 30×50 million dollars or \$1.5 billion in total. The bank can either offer the client a best efforts arrangement where it charges a fee of \$0.30 per share sold so that, assuming all shares are sold, it obtains a total fee of $0.3 \times 50 = \$15$ million. Alternatively, it can offer a firm commitment where it agrees to buy the shares from ABC Corporation for \$30 per share.

The bank is confident that it will be able to sell the shares, but is uncertain about the price. As part of its procedures for assessing risk, it considers two alternative scenarios. Under the first scenario, it can obtain a price of \$32 per share; under the second scenario, it is able to obtain only \$29 per share.

In a best-efforts deal, the bank obtains a fee of \$15 million in both cases. In a firm commitment deal, its profit depends on the price it is able to obtain. If it sells the shares for \$32, it makes a profit of $(32 - 30) \times 50 = \$100$ million because it has agreed to pay ABC Corporation \$30 per share. However, if it can only sell the shares for \$29 per share, it loses $(30 - 29) \times 50 = \$50$ million because it still has to pay ABC Corporation \$30 per share. The situation is summarized in the table below. The decision taken is likely to depend on the probabilities assigned by the bank to different outcomes and what is referred to as its “risk appetite” (see Chapter 24).

	Profits If Best Efforts	Profits If Firm Commitment
Can sell at \$29	+\$15 million	−\$50 million
Can sell at \$32	+\$15 million	+\$100 million

When equity financing is being raised and the company is already publicly traded, the investment bank can look at the prices at which the company's shares are trading a few days before the issue is to be sold as a guide to the issue price. Typically it will agree to attempt to issue new shares at a target price slightly below the current price. The main risk then is that the price of the company's shares will show a substantial decline before the new shares are sold.

2.4.1 *IPOs*

When the company wishing to issue shares is not publicly traded, the share issue is known as an *initial public offering* (IPO). This type of offering is typically made on a best efforts basis. The correct offering price is difficult to determine and depends on the investment bank's assessment of the company's value. The bank's best estimate of the market price is its estimate of the company's value divided by the number of shares currently outstanding. However, the bank will typically set the offering price below its best estimate of the market price. This is because it does not want to take the chance that the issue will not sell. (It typically earns the same fee per share sold regardless of the offering price.)

Often there is a substantial increase in the share price immediately after shares are sold in an IPO (sometimes as much as 40%), indicating that the company could have raised more money if the issue price had been higher. As a result, IPOs are considered attractive buys by many investors. Banks frequently offer IPOs to the fund managers who are their best customers and to senior executives of large companies in the hope that they will provide them with business. (The latter is known as "spinning" and is frowned upon by regulators.)

2.4.2 *Dutch Auction Approach*

Occasionally companies have used a Dutch auction approach for their IPOs. As for a regular IPO, a prospectus is issued and usually there is a road show. Individuals and companies bid by indicating the number of shares they want and the price they are prepared to pay. Shares are first issued to the highest bidder, then to the next highest bidder, and so on, until all the shares have been sold. The price paid by all successful bidders is the lowest bid that leads to a share allocation. This is illustrated in Example 2.2.

Example 2.2

A company wants to sell one million shares in an IPO. It decides to use the Dutch auction approach. The bidders are shown in the table below. In this case, shares are allocated first to C, then to F, then to E, then to H, then to A. At this point, 800,000 shares have been allocated. The next highest bidder is D, who has bid for 300,000 shares. Because only 200,000 shares remain unallocated, D's order is only two-thirds filled. The price paid by all the investors to whom shares are allocated (A, C, D, E, F, and H) is the price bid by D, or \$29.00.

Bidder	Number of Shares	Price
A	100,000	\$30.00
B	200,000	\$28.00
C	50,000	\$33.00
D	300,000	\$29.00
E	150,000	\$30.50
F	300,000	\$31.50
G	400,000	\$25.00
H	200,000	\$30.25

Dutch auctions potentially overcome two of the problems with a traditional IPO that we have mentioned. First, the price that clears the market (\$29.00 in Example 2.2) should be the market price if all potential investors have participated in the bidding process. Second, the situations where investment banks offer IPOs only to their favored clients are avoided. However, the company does not take advantage of the relationships that investment bankers have developed with large investors that usually enable the investment bankers to sell an IPO very quickly. One high-profile IPO that used a Dutch auction was the Google IPO in 2004. This is discussed in Business Snapshot 2.1.

2.4.3 SPACs

Special Purpose Acquisition Companies (SPACs) are yet another way companies can go public. A SPAC is a shell company that raises money through an IPO in order to merge with another company and thereby take it public. It typically has two years to complete an acquisition or return the funds to investors. There was a “SPAC bubble” in the first quarter of 2021. \$96 billion was raised by 295 SPACs. (As a comparison only two SPACs were created in 2010.) Business Snapshot 2.2 describes how a SPAC was used by Donald Trump.

2.4.4 Advisory Services

In addition to assisting companies with new issues of securities, investment banks offer advice to companies on mergers and acquisitions, divestments, major corporate restructurings, and so on. They will assist in finding merger partners and takeover targets or help companies find buyers for divisions or subsidiaries of which they want to divest themselves. They will also advise the management of companies that are themselves merger or takeover targets.

BUSINESS SNAPSHOT 2.1

Google's IPO

Google, developer of the well-known Internet search engine, decided to go public in 2004. It chose the Dutch auction approach. It was assisted by two investment banks, Morgan Stanley and Credit Suisse First Boston. The SEC gave approval for it to raise funds up to a maximum of \$2,718,281,828. (Why the odd number? The mathematical constant e is 2.7182818 ...) The IPO method was not a pure Dutch auction because Google reserved the right to change the number of shares that would be issued and the percentage allocated to each bidder when it saw the bids.

Some investors expected the price of the shares to be as high as \$120. But when Google saw the bids, it decided that the number of shares offered would be 19,605,052 at a price of \$85. This meant that the total value of the offering was $19,605,052 \times 85$ or \$1.67 billion. Investors who had bid \$85 or above obtained 74.2% of the shares they had bid for. The date of the IPO was August 19, 2004. Most companies would have given investors who bid \$85 or more 100% of the amount they bid for and raised \$2.25 billion, instead of \$1.67 billion. Perhaps Google (stock symbol: GOOG) correctly anticipated it would have no difficulty in selling further shares at a higher price later.

The initial market capitalization was \$23.1 billion with over 90% of the shares being held by employees. These employees included the founders, Sergey Brin and Larry Page, and the CEO, Eric Schmidt. On the first day of trading, the shares closed at \$100.34, 18% above the offer price, and there was a further 7% increase on the second day. Google's issue therefore proved to be underpriced. This may have disappointed Google's senior management as the main advantage of a Dutch-auction IPO was thought to be the avoidance of large post-IPO share price increases.

The cost of Google's IPO (fees paid to investment banks, etc.) was 2.8% of the amount raised. This compares with an average of about 4% for a regular IPO.

There were some mistakes made and Google was lucky that these did not prevent the IPO from going ahead as planned. Sergey Brin and Larry Page gave an interview to *Playboy* magazine in April 2004. The interview appeared in the September issue. This violated SEC requirements that there be a "quiet period" with no promoting of the company's stock in the period leading up to an IPO. To avoid SEC sanctions, Google had to include the *Playboy* interview (together with some factual corrections) in its SEC filings. Google also forgot to register 23.2 million shares and 5.6 million stock options.

Google's stock price rose rapidly in the period after the IPO. Approximately one year later (in September 2005) it was able to raise a further \$4.18 billion by issuing an additional 14,159,265 shares at \$295. (Why the odd number? The mathematical constant π is 3.14159265 ...)

BUSINESS SNAPSHOT 2.2

Donald Trump and a SPAC

SPACs were in the news in 2021–22 when it was revealed that a SPAC, Digital World Acquisition Corp (DWAC), was used to try and take Donald Trump's company, Media and Technology Group (TMTG), public. TMTG was created to launch a new social media platform to compete with Twitter and Facebook, both of which had banned the former president because of his role in the January 6, 2021, capitol riots. DWAC attracted the attention of regulators because it is alleged that it held meetings with Trump before going public and did not disclose this to investors.

Sometimes banks suggest steps their clients should take to avoid a merger or takeover. These are known as *poison pills*. Examples of poison pills are:

1. A potential target adds to its charter a provision where, if another company acquires one-third of the shares, other shareholders have the right to sell their shares to that company for twice the recent average share price.
2. A potential target grants to its key employees stock options that vest (i.e., can be exercised) in the event of a takeover. This makes it more difficult for the acquirer to obtain a controlling interest and is liable to create an exodus of key employees immediately after a takeover, leaving an empty shell for the new owner.
3. A potential target adds to its charter provisions making it impossible for a new owner to get rid of existing directors for one or two years after an acquisition.
4. A potential target issues preferred shares that automatically get converted to regular shares when there is a change in control.
5. A potential target adds a provision where existing shareholders have the right to purchase shares at a discounted price during or after a takeover. (This is referred to as a *flip-in*.)
6. A potential target changes the voting structure so that shares owned by management have more votes than those owned by others.

Poison pills, which are illegal in many countries outside the United States, have to be approved by a majority of shareholders. Often shareholders oppose poison pills because they see them as benefiting only management. However, they can lead to a company obtaining a higher price for its shares than would otherwise be the case.²

Valuation, strategy, and tactics are key aspects of the advisory services offered by an investment bank. For example, in advising Company A on a potential takeover of Company B, it is necessary for the investment bank to value Company B and help

²For a discussion of poison pills, see R. A. Heron and E. Lie. "On the use of poison pills and defensive payouts by takeover targets." *Journal of Business*, 79, no. 4 (2006): 1783–1807.

Company A assess possible synergies between the operations of the two companies. It must also consider whether it is better to offer Company B's shareholders cash or a share-for-share exchange (i.e., a certain number of shares in Company A in exchange for each share of Company B). What should the initial offer be? What does it expect the final offer that will close the deal to be? It must assess the best way to approach the senior managers of Company B and consider what the motivations of the managers will be. Will the takeover be a hostile one (opposed by the management of Company B) or friendly one (supported by the management of Company B)? In some instances there will be antitrust issues, and approval from some branch of government may be required.

2.5 Securities Trading

Banks often get involved in securities trading, providing brokerage services, and making a market in individual securities. In doing so, they compete with smaller securities firms that do not offer other banking services. The Dodd-Frank Act in the United States, which will be discussed later in this book, does not allow banks to engage in proprietary trading. In some other countries, proprietary trading is allowed, but it usually has to be organized so that losses do not affect depositors.

Most large investment and commercial banks have extensive trading activities. Apart from proprietary trading (which may or may not be allowed), banks trade to provide services to their clients. (For example, a bank might enter into a derivatives transaction with a corporate client to help it reduce its foreign exchange risk.) They also trade (typically with other financial institutions) to hedge their risks.

A broker assists in the trading of securities by taking orders from clients and arranging for them to be carried out on an exchange. Some brokers operate nationally, and some serve only a particular region. Some, known as full-service brokers, offer investment research and advice. Others, known as discount brokers, charge lower commissions, but provide no advice. Some offer online services, and some, such as E*Trade, provide a platform for customers to trade without a broker.

A market maker facilitates trading by always being prepared to quote a bid (the price at which it is prepared to buy) and an ask (the price at which it is prepared to sell, also known as an offer price). When providing a quote, it does not know whether the person requesting the quote wants to buy or sell. The market maker makes a profit from the spread between the bid and the ask, but takes the risk that it will be left with a big long or short position and lose money.

Many exchanges on which stocks, options, and futures trade use market makers. Typically, an exchange will specify a maximum level for the size of a market maker's bid-ask spread (the difference between the ask and the bid). Banks are often market makers for instruments such as forward contracts, swaps, and options. (See Chapter 5 for a discussion of these instruments.)

2.6 Potential Conflicts of Interest in Banking

There are many potential conflicts of interest between commercial banking, securities services, and investment banking when they are all conducted under the same corporate umbrella. For example:

1. When asked for advice by an investor, a bank might be tempted to recommend securities that the investment banking part of its organization is trying to sell. When it has a fiduciary account (i.e., a customer account where the bank can choose trades for the customer), the bank can “stuff” difficult-to-sell securities into the account.
2. A bank, when it lends money to a company, often obtains confidential information about the company. It might be tempted to pass that information to the mergers and acquisitions arm of the investment bank to help it provide advice to one of its clients on potential takeover opportunities.
3. The research end of the securities business might be tempted to recommend a company’s share as a “buy” in order to please the company’s management and obtain investment banking business.
4. Suppose a commercial bank no longer wants a loan it has made to a company on its books because the confidential information it has obtained from the company leads it to believe that there is an increased chance of bankruptcy. It might be tempted to ask the investment bank to arrange a bond issue for the company, with the proceeds being used to pay off the loan. This would have the effect of replacing its loan with a loan made by investors who were less well informed.

As a result of these types of conflicts of interest, some countries have in the past attempted to separate commercial banking from investment banking. The Glass-Steagall Act of 1933 in the United States limited the ability of commercial banks and investment banks to engage in each other’s activities. Commercial banks were allowed to continue underwriting Treasury instruments and some municipal bonds. They were also allowed to do private placements. But they were not allowed to engage in other activities such as public offerings. Similarly, investment banks were not allowed to take deposits and make commercial loans.

In 1987, the Federal Reserve Board relaxed the rules somewhat and allowed banks to establish holding companies with two subsidiaries, one in investment banking and the other in commercial banking. The revenue of the investment banking subsidiary was restricted to being a certain percentage of the group’s total revenue.

In 1997, the rules were relaxed further so that commercial banks could acquire existing investment banks. Finally, in 1999, the Financial Services Modernization Act was passed. This effectively eliminated all restrictions on the operations of banks, insurance companies, and securities firms. In 2007, there were five large investment banks in the United States that had little or no commercial banking interests. These were Goldman Sachs, Morgan Stanley, Merrill Lynch, Bear Stearns, and Lehman Brothers. In 2008, the Global Financial Crisis led to Lehman Brothers going bankrupt, Bear Stearns being taken over by JPMorgan Chase, and Merrill Lynch being taken over by

Bank of America. Goldman Sachs and Morgan Stanley became bank holding companies with both commercial and investment banking interests. (As a result, they have had to subject themselves to more regulatory scrutiny.) The year 2008 therefore marked the end of an era for investment banking in the United States.

We have not returned to the Glass–Steagall world where investment banks and commercial banks were kept separate. But increasingly banks are required to ring-fence their deposit-taking businesses so that they cannot be affected by losses in investment banking.

2.7 Today's Large Banks

Today's large banks operate globally and transact business in many different areas. They are still engaged in the traditional commercial banking activities of taking deposits, making loans, and clearing checks (both nationally and internationally). They offer retail customers credit cards, telephone banking, Internet banking, and automatic teller machines (ATMs). They provide payroll services to businesses and, as already mentioned, they have large trading activities.

Banks offer lines of credit to businesses and individual customers. They provide a range of services to companies when they export goods and services. Companies can enter into a variety of contracts with banks that are designed to hedge risks they face relating to foreign exchange, commodity prices, interest rates, and other market variables. These contracts will be discussed in later chapters. Even risks related to the weather can be hedged.

Banks undertake securities research and offer “buy,” “sell,” and “hold” recommendations on individual stocks. They offer brokerage services (discount and full service). They offer trust services where they are prepared to manage portfolios of assets for clients. They have economics departments that consider macroeconomic trends and actions likely to be taken by central banks. These departments produce forecasts on interest rates, exchange rates, commodity prices, and other variables. Banks offer a range of mutual funds and in some cases have their own hedge funds. Increasingly banks are offering insurance products.

The investment banking arm of a bank has complete freedom to underwrite securities for governments and corporations. It can provide advice to corporations on mergers and acquisitions and other topics relating to corporate finance.

How are the conflicts of interest outlined in Section 2.6 handled? There are internal barriers known as *Chinese walls*. These internal barriers prohibit the transfer of information from one part of the bank to another when this is not in the best interests of one or more of the bank's customers. There have been some well-publicized violations of conflict-of-interest rules by large banks. These have led to hefty fines and lawsuits. Top management has a big incentive to enforce Chinese walls. This is not only because of the fines and lawsuits. A bank's reputation is its most valuable asset. The adverse publicity associated with conflict-of-interest violations can lead to a loss of confidence in the bank and business being lost in many different areas.

2.7.1 *Accounting*

It is appropriate at this point to provide a brief discussion of how a bank calculates a profit or loss from its many diverse activities. Activities that generate fees, such as most investment banking activities, are straightforward. Accrual accounting rules similar to those that would be used by any other business apply.

For other banking activities, there is an important distinction between the “banking book” and the “trading book.” As its name implies, the trading book includes all the assets and liabilities the bank has as a result of its trading operations. The values of these assets and liabilities are *marked to market* daily. This means that the value of the book is adjusted daily to reflect changes in market prices. If a bank trader buys an asset for \$100 on one day and the price falls to \$60 the next day, the bank records an immediate loss of \$40—even if it has not sold the asset. Sometimes it is not easy to estimate the value of a contract that has been entered into because there are no market prices for similar transactions. For example, there might be a lack of liquidity in the market or it might be the case that the transaction is a complex nonstandard derivative that does not trade sufficiently frequently for benchmark market prices to be available. Banks are nevertheless expected to come up with a market price in these circumstances. Often a model has to be assumed. The process of coming up with a “market price” is then sometimes termed *marking to model*. (Chapter 22 discusses model risk and accounting issues further.)

The banking book includes loans made to corporations and individuals. Traditionally these have not been marked to market. However, this has changed as a result of a new accounting standard from the International Accounting Standards Board, IFRS 9, and similar accounting standards from the Financial Accounting Standards Board (FASB) in the United States. The new accounting rules require lenders to estimate the expected credit losses from their loan portfolios and adjust the value of the loan portfolios accordingly. The period of time over which expected losses are estimated is one year for IFRS 9 and the loan’s lifetime for FASB, but when the credit risk is large IFRS 9 also calculates expected losses over the lifetime. The new rules are partly a result of the Global Financial Crisis. Banks had portfolios of subprime mortgages that performed badly. However, losses did not have to be reported until they had been actually incurred.

Occasionally banks have resorted to artificial ways of avoiding the recognition of loan losses, as outlined in Business Snapshot 2.3.

2.7.2 *The Originate-to-Distribute Model*

DLC, the small hypothetical bank we looked at in Tables 2.2 to 2.4, took deposits and used them to finance loans. An alternative approach is known as the *originate-to-distribute model*. This involves the bank originating but not keeping loans. Portfolios of loans are packaged into tranches that are then sold to investors.

BUSINESS SNAPSHOT 2.3

How to Avoid Loan Losses

When a borrower is experiencing financial difficulties and is unable to make interest and principal payments as they become due, it is sometimes tempting to lend more money to the borrower so that the payments on the old loans can be kept up to date. No losses are then recorded.

In the 1970s, banks in the United States and other countries lent huge amounts of money to Eastern European, Latin American, and other less developed countries (LDCs). Some of the loans were made to help countries develop their infrastructure, but others were less justifiable (e.g., one was to finance the coronation of a ruler in Africa). Sometimes the money found its way into the pockets of dictators. For example, the Marcos family in the Philippines allegedly transferred billions of dollars into its own bank accounts.

In the early 1980s, many LDCs were unable to service their loans. One option for them was *debt repudiation*, but a more attractive alternative was *debt rescheduling*, where more money was lent to cover required payments on existing loans.

In 1987, Citicorp (now Citigroup) took the lead in refusing to reschedule LDC debt and increased its loan loss reserves by \$3 billion in recognition of expected losses on the debt. Other banks with large LDC exposures followed suit.

The originate-to-distribute model has been used in the U.S. mortgage market for many years. In order to increase the liquidity of the U.S. mortgage market and facilitate the growth of home ownership, three government-sponsored entities have been created: the Government National Mortgage Association (GNMA) or “Ginnie Mae,” the Federal National Mortgage Association (FNMA) or “Fannie Mae,” and the Federal Home Loan Mortgage Corporation (FHLMC) or “Freddie Mac.” These agencies buy pools of mortgages from banks and other mortgage originators, guarantee the timely repayment of interest and principal, and then package the cash flow streams and sell them to investors. The investors typically take what is known as prepayment risk. This is the risk that interest rates will decrease and mortgages will be paid off earlier than expected. However, they do not take any credit risk because the mortgages are guaranteed by GNMA, FNMA, or FHLMC. In 1999, the agencies started to guarantee subprime loans and as a result ran into serious financial difficulties.³

³GNMA has always been government owned whereas FNMA and FHLMC used to be private corporations with shareholders. As a result of their financial difficulties in 2008, the U.S. government had to step in and assume complete control of FNMA and FHLMC.

The originate-to-distribute model has been used for many types of bank lending, including student loans, commercial loans, commercial mortgages, residential mortgages, and credit card receivables. In many cases there is no guarantee that payment will be made so that it is the investors who bear the credit risk when the loans are packaged and sold.

The originate-to-distribute model is also termed *securitization* because securities are created from cash flow streams originated by the bank. It is an attractive model for banks. By securitizing its loans it gets them off the balance sheet and frees up funds to enable it to make more loans. It also frees up capital that can be used to cover risks being taken elsewhere in the bank. (This is particularly attractive if the bank feels that the capital required by regulators for a loan is too high.) A bank earns a fee for originating a loan and a further fee if it services the loan after it has been sold.

As we will explain in Chapter 7, the originate-to-distribute model got out of control during the 2000 to 2006 period. Banks relaxed their mortgage lending standards and the credit quality of the instruments being originated declined sharply. This led to a severe financial crisis and a period during which the originate-to-distribute model could not be used by banks because investors had lost confidence in the securities that had been created.

2.8 The Risks Facing Banks

A bank's operations give rise to many risks. Much of the rest of this book is devoted to considering these risks in detail.

Central bank regulators require banks to hold capital for the risks they are bearing. In 1988, international standards were developed for the determination of this capital. These standards and the way they have evolved since 1988 are discussed in Chapters 25, 26, and 27. Capital is now required for three types of risk: credit risk, market risk, and operational risk.

Credit risk is the risk that counterparties in loan transactions and derivatives transactions will default. This has traditionally been the greatest risk facing a bank. Market risk arises primarily from the bank's trading operations. It is the risk relating to the possibility that instruments in the bank's trading book will decline in value. Operational risk, which is now often considered to be the biggest risk facing banks, is the risk that losses are created because internal systems fail to work as they are supposed to or because of external events. The time horizon used by regulators for considering losses from credit risks and operational risks is one year, whereas the time horizon for considering losses from market risks is usually much shorter. The objective of regulators is to keep the total capital of a bank sufficiently high that the chance of a bank failure is very low.

In addition to calculating regulatory capital, most large banks have systems in place for calculating what is termed *economic capital* (see Chapter 28). This is the capital that the bank, using its own models rather than those prescribed by regulators, thinks it needs. Economic capital is often less than regulatory capital. However, banks have no choice but to maintain their capital above the regulatory capital level. The form the capital can take (equity, subordinated debt, etc.) is prescribed by regulators. To avoid having to raise

capital at short notice, banks try to keep their capital comfortably above the regulatory minimum.

When banks announced huge losses on their subprime mortgage portfolios in 2007 and 2008, many had to raise new equity capital in a hurry. *Sovereign wealth funds*, which are investment funds controlled by the government of a country, have provided some of this capital. For example, Citigroup, which reported losses in the region of \$40 billion, raised \$7.5 billion in equity from the Abu Dhabi Investment Authority in November 2007 and \$14.5 billion from investors that included the governments of Singapore and Kuwait in January 2008. Later, Citigroup and many other banks required capital injections from their own governments to survive.

Summary

Banks are complex global organizations engaged in many different types of activities. Today, the world's large banks are engaged in taking deposits, making loans, underwriting securities, trading, providing brokerage services, providing fiduciary services, advising on a range of corporate finance issues, offering mutual funds, providing services to hedge funds, and so on. There are potential conflicts of interest and banks develop internal rules to avoid them. It is important that senior managers are vigilant in ensuring that employees obey these rules. The cost in terms of reputation, lawsuits, and fines from inappropriate behavior where one client (or the bank) is advantaged at the expense of another client can be very high.

There are now international agreements on the regulation of banks. This means that the capital banks are required to keep for the risks they are bearing does not vary too much from one country to another. Many countries have guaranty programs that protect small depositors from losses arising from bank failures. This has the effect of maintaining confidence in the banking system and avoiding mass withdrawals of deposits when there is negative news (or perhaps just a rumor) about problems faced by a particular bank.

Further Reading

Heron, R. A. and E. Lie. "On the use of poison pills and defensive payouts by takeover targets." *Journal of Business*, 79, no. 4 (2006): 1783–1807.

Practice Questions and Problems (Answers at End of Book)

- 2.1 How did concentration in the U.S. banking system change between 1984 and 2022?
- 2.2 What government policies led to the large number of small community banks in the United States?

- 2.3 What risks does a bank take if it funds long-term loans with short-term deposits?
- 2.4 Suppose that an out-of-control trader working for DLC Bank (see Tables 2.2 and 2.3) loses \$7 million trading foreign exchange. What do you think would happen?
- 2.5 What is meant by net interest income?
- 2.6 Which items on the income statement of DLC Bank in Section 2.2 are most likely to be affected by (a) credit risk, (b) market risk, and (c) operational risk?
- 2.7 Explain the terms “private placement” and “public offering.” What is the difference between “best efforts” and “firm commitment” for a public offering?
- 2.8 The bidders in a Dutch auction are as follows:

Bidder	Number of Shares	Price
A	20,000	\$100.00
B	30,000	\$93.00
C	50,000	\$110.00
D	70,000	\$88.00
E	60,000	\$80.00
F	10,000	\$105.00
G	90,000	\$70.00
H	80,000	\$125.00

The number of shares being auctioned is 150,000. What is the price paid by investors? How many shares does each investor receive?

- 2.9 What is the attraction of a Dutch auction over the normal procedure for an IPO? In what ways was Google’s IPO different from a standard Dutch auction?
- 2.10 Management sometimes argues that poison pills are in the best interests of shareholders because they enable management to extract a higher price from would-be acquirers. Discuss this argument.
- 2.11 Give three examples of the conflicts of interest in a large bank. How are conflicts of interest handled?
- 2.12 What is the difference between the banking book and the trading book?
- 2.13 How has accounting for loans changed since the 2007–2008 Global Financial Crisis?
- 2.14 What is the originate-to-distribute model?

Further Questions

- 2.15 Regulators calculate that DLC bank (see Section 2.2) will report a profit that is normally distributed with a mean of \$0.6 million and a standard deviation of \$2 million. How much equity capital in addition to that in Table 2.2 should regulators require for there to be a 99.9% chance of the capital not being wiped out by losses?

- 2.16 Explain the moral hazard problems with deposit insurance. How can they be overcome?
- 2.17 The bidders in a Dutch auction are as follows:

Bidder	Number of Shares	Price
A	60,000	\$50.00
B	20,000	\$80.00
C	30,000	\$55.00
D	40,000	\$38.00
E	40,000	\$42.00
F	40,000	\$42.00
G	50,000	\$35.00
H	50,000	\$60.00

The number of shares being auctioned is 210,000. What is the price paid by investors? How many shares does each investor receive?

- 2.18 An investment bank has been asked to underwrite an issue of 10 million shares by a company. It is trying to decide between a firm commitment where it buys the shares for \$10 per share and a best efforts where it charges a fee of 20 cents for each share sold. Explain the pros and cons of the two alternatives.

Chapter 3

Insurance Companies and Pension Plans

The role of insurance companies is to provide protection against adverse events. The company or individual seeking protection is referred to as the *policyholder*. The policyholder makes regular payments, known as *premiums*, and receives payments from the insurance company if certain specified events occur. Insurance is usually classified as *life insurance* and *nonlife insurance*, with health insurance often considered to be a separate category. Nonlife insurance is also referred to as *property-casualty insurance*, and this is the terminology we will use here.

A life insurance contract typically lasts a long time and provides payments to the policyholder's beneficiaries that depend on when the policyholder dies. A property-casualty insurance contract typically lasts one year (although it may be renewed) and provides compensation for losses from accidents, fire, theft, and so on.

Insurance has existed for many years. As long ago as 200 B.C., there was an arrangement in ancient Greece where an individual could make a lump-sum payment (the amount dependent on his or her age) and obtain a monthly income for life. The Romans had a form of life insurance where an individual could purchase a contract that would provide a payment to relatives on his or her death. In ancient China, a form of property-casualty insurance existed between merchants where, if the ship of one merchant sank, the rest of the merchants would provide compensation.

An employer pension plan is a form of insurance arranged by a company for its employees. It is designed to provide the employees with income for the rest of their

lives once they have retired. Typically both the company and its employees make regular monthly contributions to the plan and the funds in the plan are invested to provide income for retirees.

This chapter describes how the contracts offered by insurance companies work. It explains the risks that insurance companies face and the way they are regulated. It also discusses key issues associated with pension plans.

3.1 Life Insurance

In life insurance contracts, the payments to the policyholder depend—at least to some extent—on when the policyholder dies. Outside the United States, the term *life assurance* is often used to describe a contract where the event being insured against is certain to happen at some future time (e.g., a contract that will pay \$100,000 on the policyholder's death). Life insurance is used to describe a contract where the event being insured against may never happen (for example, a contract that provides a payoff in the event of the accidental death of the policyholder). In the United States, all types of life policies are referred to as life insurance and this is the terminology that will be adopted here.

There are many different types of life insurance products. The products available vary from country to country. We will now describe some of the more common ones.

3.1.1 Term Life Insurance

Term life insurance (sometimes referred to as *temporary life insurance*) lasts a predetermined number of years. If the policyholder dies during the life of the policy, the insurance company makes a payment to the specified beneficiaries equal to the face amount of the policy. If the policyholder does not die during the term of the policy, no payments are made by the insurance company. The policyholder is required to make regular monthly or annual premium payments (usually constant predetermined amounts) to the insurance company for the life of the policy or until the policyholder's death (whichever is earlier). The face amount of the policy typically stays the same or declines with the passage of time. A policy where the premium per year is not constant is an *annual renewable term* policy. In this, the insurance company guarantees to renew the policy from one year to the next at a rate reflecting the policyholder's age without regard to the policyholder's health.

A common reason for term life insurance is a mortgage. For example, a person age 35 with a 25-year mortgage might choose to buy 25-year term insurance (with a declining face amount) to provide dependents with the funds to pay off the mortgage in the event of his or her death.

3.1.2 Whole Life Insurance

Whole life insurance (sometimes referred to as *permanent life insurance*) provides protection for the life of the policyholder. The policyholder is required to make regular monthly or annual payments until his or her death. The face value of the policy is then paid to the

designated beneficiary. In the case of term life insurance, there is no certainty that there will be a payout, but in the case of whole life insurance, a payout is certain to happen providing the policyholder continues to make the agreed premium payments. The only uncertainty is when the payout will occur. Not surprisingly, whole life insurance requires considerably higher premiums than term life insurance policies. Usually, the payments and the face value of the policy both remain constant through time.

Policyholders can often redeem (surrender) whole life policies early or use the policies as collateral for loans. When a policyholder wants to redeem a whole life policy early, it is sometimes the case that an investor will buy the policy from the policyholder for more than the surrender value offered by the insurance company. The investor will then make the premium payments and collect the face value from the insurance company when the policyholder dies.

Suppose a man buys \$1 million of whole life insurance at age 40 and that the premium is \$15,000 per year. In the first year, the policy will on average cost the insurance company much less than \$15,000. As we will see later, an estimate of the probability of a male aged 40 dying within one year is 0.002581. The expected cost to the insurance company in the first year is therefore $\$1,000,000 \times 0.002581$ or \$2,581. There is therefore an expected cash inflow of \$12,419 to the insurance company in the first year. Assuming that the policy holder does not die in the first year, a premium of \$15,000 is paid at the beginning of the second year. The probability of a man aged 41 dying within one year is 0.002697. This means that there is an expected cash inflow of \$12,303 in the second year. The expected costs of one-year premiums continue to rise as the individual gets older and at some stage the expected cost of the policy in a year becomes greater than \$15,000. For example, assuming the man has not already died, he will pay a premium of \$15,000 when he reaches 70 and have a probability of 0.022364 of dying within a year. This means that there is an expected cash outflow to the insurance company of \$22,364 less \$15,000 or \$7,364. This expected cash outflow, as seen at the beginning of the life of the policy, is \$7,364 times the probability that a man aged 40 is still alive after 30 years, which turns out to be \$5,634. The year-by-year (undiscounted) expected cash flows, estimated at beginning of the policy, are shown in Figure 3.1. The insurance company has expected cash inflows, which we will refer to as *surplus premiums*, for the first 24 years. After that the expected cash flow in a year is negative. (Note that by the 60th year of the policy the expected cash flow is close to zero because the probability that the policy holder is still alive then is small.) Figure 3.1 shows that there is a savings element to whole life insurance. In the early years, expected surpluses can be invested to cover later expected payouts.

There are tax advantages associated with life insurance policies in many countries. If the policyholder invested the surplus premiums instead of paying them to the insurance company, tax would normally be payable on the income as it was earned. But, when the surplus premiums are invested within the insurance policy, the tax treatment is often better. Tax is deferred, and sometimes the payout to the beneficiaries of life insurance policies is free of income tax altogether.

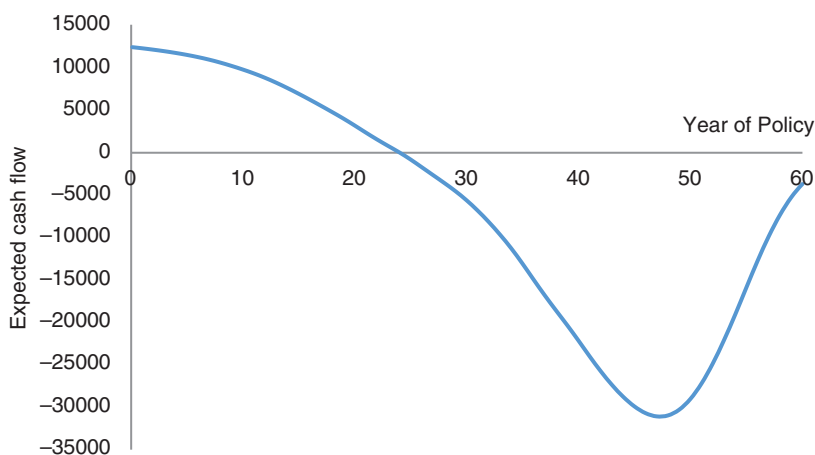


Figure 3.1 Expected Cash Flows to Insurance Company from a Whole Life Insurance Policy Taken Out by a Man Aged 40

3.1.3 Variable Life Insurance

Given that a whole life insurance policy involves funds being invested for the policyholder, a natural development is to allow the policyholder to specify how the funds are invested. Variable life (VL) insurance is a form of whole life insurance where the surplus premiums discussed earlier are invested in a fund chosen by the policyholder. This could be an equity fund, a bond fund, or a money market fund. A minimum guaranteed payout on death is usually specified, but the payout can be more if the fund does well. Income earned from the investments can sometimes be applied toward the premiums. The policyholder can usually switch from one fund to another at any time.

3.1.4 Universal Life

Universal life (UL) insurance is also a form of whole life insurance. The policyholder can reduce the premium down to a specified minimum without the policy lapsing. The surplus premiums are invested by the insurance company in fixed income products such as bonds, mortgages, and money market instruments. The insurance company guarantees a certain minimum return, say 4%, on these funds. The policyholder can choose between two options. Under the first option, a fixed benefit is paid on death; under the second option, the policyholder's beneficiaries receive more than the fixed benefit if the investment return is greater than the guaranteed minimum. Needless to say, premiums are lower for the first option.

3.1.5 Variable-Universal Life Insurance

Variable-universal life (VUL) insurance blends the features found in variable life insurance and universal life insurance. The policyholder can choose between a number

of alternatives for the investment of surplus premiums. The insurance company guarantees a certain minimum death benefit and interest on the investments can sometimes be applied toward premiums. Premiums can be reduced down to a specified minimum without the policy lapsing.

3.1.6 Endowment Life Insurance

Endowment life insurance lasts for a specified period and pays a lump sum either when the policyholder dies or at the end of the period, whichever is first. There are many different types of endowment life insurance contracts. The amount that is paid out can be specified in advance as the same regardless of whether the policyholder dies or survives to the end of the policy. Sometimes the payout is also made if the policyholder has a critical illness. In a *with-profits* endowment life insurance policy, the insurance company declares periodic bonuses that depend on the performance of the insurance company's investments. These bonuses accumulate to increase the amount paid out to the policyholder, assuming the policyholder lives beyond the end of the life of the policy. In a unit-linked endowment, the amount paid out at maturity depends on the performance of the fund chosen by the policyholder. A *pure endowment* policy has the property that a payout occurs only if the policyholder survives to the end of the life of the policy.

3.1.7 Group Life Insurance

Group life insurance covers many people under a single policy. It is often purchased by a company for its employees. The policy may be *contributory*, where the premium payments are shared by the employer and employee, or *noncontributory*, where the employer pays the whole of the cost. There are economies of scale in group life insurance. The selling and administration costs are lower. An individual is usually required to undergo medical tests when purchasing life insurance in the usual way, but this may not be necessary for group life insurance. The insurance company knows that it will be taking on some better-than-average risks and some worse-than-average risks.

3.2 Annuity Contracts

Many life insurance companies also offer annuity contracts. While a life insurance contract has the effect of converting regular payments into a lump sum, an annuity contract has the opposite effect: that of converting a lump sum into regular payments. In a typical arrangement, the policyholder makes a lump-sum payment to the insurance company and the insurance company agrees to provide the policyholder with an annuity that starts at a particular date and lasts for the rest of the policyholder's life. In some instances, the annuity starts immediately after the lump-sum payment by the policyholder. In other cases, the lump-sum payment is made by the policyholder several years ahead of the

time when the annuity is to start and the insurance company invests the funds to create the annuity. (This is referred to as a *deferred annuity*.) Instead of a lump sum, the policyholder sometimes saves for the annuity by making regular monthly, quarterly, or annual payments to the insurance company.

There are often tax deferral advantages to the policyholder. This is because taxes usually have to be paid only when the annuity income is received. The amount to which the funds invested by the insurance company on behalf of the policyholder have grown in value is sometimes referred to as the *accumulation value*. Funds can usually be withdrawn early, but there are liable to be penalties. In other words, the surrender value of an annuity contract is typically less than the accumulation value. This is because the insurance company has to recover selling and administration costs. Policies sometimes allow *penalty-free withdrawals* where a certain percentage of the accumulation value or a certain percentage of the original investment can be withdrawn in a year without penalty. In the event that the policyholder dies before the start of the annuity (and sometimes in other circumstances such as when the policyholder is admitted to a nursing home), the full accumulation value can often be withdrawn without penalty.

Some deferred annuity contracts in the United States have embedded options. The accumulation value is sometimes calculated so that it tracks a particular equity index, such as the S&P 500. Lower and upper limits are specified. If the growth in the index in a year is less than the lower limit, the accumulation value grows at the lower limit rate; if it is greater than the upper limit, the accumulation value grows at the upper limit rate; otherwise it grows at the same rate as the S&P 500. Suppose that the lower limit is 0% and the upper limit is 8%. The policyholder is assured that the accumulation value will never decline, but index growth rates in excess of 8% are given up. In this type of arrangement, the policyholder is typically not compensated for dividends that would be received from an investment in the stocks underlying the index and the insurance company may be able to change parameters such as the lower limit and the upper limit from one year to the next. These types of contracts appeal to investors who want an exposure to the equity market but are reluctant to risk a decline in their accumulation value. Sometimes, the way the accumulation value grows from one year to the next is a quite complicated function of the performance of the index during the year.

In the United Kingdom, the annuity contracts offered by insurance companies used to guarantee a minimum level for the interest rate used for the calculation of the size of the annuity payments. Many insurance companies regarded this guarantee—an interest rate option granted to the policyholder—as a necessary marketing cost and did not calculate the cost of the option or hedge their risks. As interest rates declined and life expectancies increased, many insurance companies found themselves in financial difficulties and, as described in Business Snapshot 3.1, at least one of them went bankrupt.

BUSINESS SNAPSHOT 3.1

Equitable Life

Equitable Life was a British life insurance company founded in 1762 that at its peak had 1.5 million policyholders. Starting in the 1950s, Equitable Life sold annuity products where it guaranteed that the interest rate used to calculate the size of the annuity payments would be above a certain level. (This is known as a Guaranteed Annuity Option, GAO.) The guaranteed interest rate was gradually increased in response to competitive pressures and increasing interest rates. Toward the end of 1993, interest rates started to fall. Also, life expectancies were rising so that the insurance companies had to make increasingly high provisions for future payouts on contracts. Equitable Life did not take action. Instead, it grew by selling new products. In 2000, it was forced to close its doors to new business. A report issued by Ann Abraham in July 2008 was highly critical of regulators and urged compensation for policyholders.

An interesting aside to this is that regulators did at one point urge insurance companies that offered GAOs to hedge their exposures to an interest rate decline. As a result, many insurance companies scrambled to enter into contracts with banks that paid off if long-term interest rates declined. The banks in turn hedged their risk by buying instruments such as bonds that increased in price when rates fell. This was done on such a massive scale that the extra demand for bonds caused long-term interest rates in the UK to decline sharply (increasing losses for insurance companies on the unhedged part of their exposures). This shows that when large numbers of different companies have similar exposures, problems are created if they all decide to hedge at the same time. There are not likely to be enough investors willing to take on their risks without market prices changing.

3.3 Mortality Tables

Mortality tables are the key to valuing life insurance contracts. Table 3.1 shows an extract from the mortality rates estimated by the U.S. Department of Social Security from statistics on death rates as a function of age collected in 2019. To understand the table, consider the row corresponding to age 31. The second column shows that the probability of a man who has just reached age 31 dying within the next year is 0.001860 (or 0.1860%). The third column shows that the probability of a man surviving to age 31 is 0.97196 (or 97.196%). The fourth column shows that a man age 31 has a remaining life expectancy

Table 3.1 Mortality Table from 2019 Data

Age (Years)	Male			Female		
	Probability of Death within 1 Year	Survival Probability	Life Expectancy	Probability of Death within 1 Year	Survival Probability	Life Expectancy
0	0.006081	1.00000	76.23	0.005046	1.00000	81.28
1	0.000425	0.99392	75.69	0.000349	0.99495	80.69
2	0.000260	0.99350	74.73	0.000212	0.99461	79.72
3	0.000194	0.99324	73.75	0.000166	0.99440	78.74
...	...	...	...	...	...	...
30	0.001798	0.97371	47.83	0.000811	0.98594	52.26
31	0.001860	0.97196	46.92	0.000872	0.98514	51.30
32	0.001926	0.97015	46.00	0.000933	0.98428	50.34
33	0.001994	0.96828	45.09	0.000990	0.98336	49.39
...	...	...	...	...	...	...
40	0.002581	0.95324	38.75	0.001423	0.97532	42.76
41	0.002697	0.95078	37.84	0.001502	0.97394	41.83
42	0.002828	0.94821	36.95	0.001596	0.97247	40.89
43	0.002976	0.94553	36.05	0.001709	0.97092	39.95
...	...	...	...	...	...	...
50	0.004888	0.92168	29.88	0.002980	0.95617	33.51
51	0.005319	0.91718	29.03	0.003240	0.95332	32.61
52	0.005808	0.91230	28.18	0.003529	0.95023	31.71
53	0.006360	0.90700	27.34	0.003852	0.94688	30.82
...	...	...	...	...	...	...
60	0.011447	0.85506	21.77	0.006880	0.91396	24.80
61	0.012352	0.84527	21.02	0.007454	0.90767	23.96
62	0.013248	0.83483	20.28	0.008006	0.90091	23.14
63	0.014117	0.82377	19.54	0.008515	0.89370	22.32
...	...	...	...	...	...	...
70	0.022364	0.72924	14.60	0.014662	0.82955	16.76
71	0.024169	0.71293	13.92	0.016210	0.81739	16.00
72	0.026249	0.69570	13.25	0.017892	0.80414	15.26
73	0.028642	0.67744	12.59	0.019701	0.78975	14.52
...	...	...	...	...	...	...
80	0.056205	0.51307	8.43	0.041102	0.64992	9.83
81	0.062327	0.48423	7.91	0.046080	0.62321	9.23
82	0.069190	0.45405	7.40	0.051658	0.59449	8.65
83	0.076844	0.42264	6.91	0.057868	0.56378	8.09
...	...	...	...	...	...	...
90	0.161582	0.18913	4.12	0.128749	0.30509	4.88
91	0.178797	0.15857	3.82	0.143532	0.26581	4.53
92	0.197287	0.13022	3.54	0.159491	0.22766	4.20
93	0.217013	0.10453	3.29	0.176600	0.19135	3.90

SOURCE: U.S. Department of Social Security, www.ssa.gov/OACT/STATS/table4c6.html.

of 46.92 years. This means that on average he will live to age 77.92. The remaining three columns show similar statistics for a woman. The probability of a 31-year-old woman dying within one year is 0.000872 (0.0872%), the probability of a woman surviving to age 31 is 0.98514 (98.514%), and the remaining life expectancy for a 31-year-old woman is 51.30 years.

The full table shows that the probability of death during the following year is a decreasing function of age for the first 10 years of life and then starts to increase. Mortality statistics for women are a little more favorable than for men. If a man is lucky enough to reach age 90, the probability of death in the next year is about 16.2%. The full table shows this probability is about 34.8% at age 100 and 56.7% at age 110. For women, the corresponding probabilities are 12.9%, 29.9%, and 53.6%, respectively.

All the numbers in the table can be calculated from the “Probability of Death within 1 Year” column. The probability that an individual will survive to age $n + 1$ is the probability that the individual will survive to age n multiplied by 1 minus the probability that an individual aged n will die within the next year. For example, the probability of surviving to age 61 can be calculated as $0.85506 \times (1 - 0.11447)$.

Next consider life expectancy for an individual of a certain age. This can be determined from the probability of death during the first year, the second year, the third year, and so on. Consider a man age 90. The probability that the man will die within one year is, from the table, 0.161582. The probability that he will die in the second year (between ages 91 and 92) is the probability that he does not die in the first year multiplied by the probability that he does die in the second year. From the numbers in the second column of the table, this is

$$(1 - 0.161582) \times 0.178797 = 0.149907$$

Similarly, the probability that he dies in the third year (between ages 92 and 93) is

$$(1 - 0.161582) \times (1 - 0.178797) \times 0.197287 = 0.135834$$

Assuming that death occurs on average halfway through a year, the life expectancy of a man age 90 is

$$0.5 \times 0.161582 + 1.5 \times 0.149907 + 2.5 \times 0.135834 + \cdots$$

Example 3.1

Assume that interest rates for all maturities are 4% per annum (with semiannual compounding) and premiums are paid once a year at the beginning of the year. What is an insurance company’s break-even premium for \$100,000 of term life insurance for a man of average health age 90? If the term insurance lasts one year, the expected payout is $0.161582 \times 100,000$ or \$16,158. Assume that the payout occurs halfway through the year. (This is likely to be approximately true on average.) The break-even premium is \$16,158 discounted for six months. This is $16,152/1.02$ or \$15,841.

Suppose next that the term insurance lasts two years. In this case, the present value of expected payout in the first year is \$15,841 as before. The probability that the policyholder dies during the second year is $(1 - 0.161582) \times 0.178797 = 0.149907$ so that there is also an expected payout of $0.149907 \times 100,000$ or \$14,991 during the second year. Assuming this happens at time 18 months, the present value of the payout is $14,991/(1.02^3)$ or \$14,126. The total present value of payouts is $15,841 + 14,126$ or \$29,967.

Consider next the premium payments. The first premium is required at time zero, so we are certain that this will be paid. The probability of the second premium payment being made at the beginning of the second year is the probability that the man does not die during the first year. This is $1 - 0.161582 = 0.838418$. When the premium is X dollars per year, the present value of the premium payments is

$$X + \frac{0.838418X}{(1.02)^2} = 1.805861X$$

The break-even annual premium is given by the value of X that equates the present value of the expected premium payments to the present value of the expected payout. This is the value of X that solves

$$1.805861X = 29,967$$

or $X = 16,594$. The break-even premium payment is therefore \$16,594.

3.4 Longevity and Mortality Risk

Longevity risk is the risk that advances in medical sciences and lifestyle changes will lead to people living longer. Increases in longevity adversely affect the profitability of most types of annuity contracts (because the annuity has to be paid for longer), but increase the profitability of most life insurance contracts (because the final payout is either delayed or, in the case of term insurance, less likely to happen). Life expectancy varies from country to country, but has been steadily increasing in most parts of the world. Average life expectancy of a child born in the United States in 2019 is estimated to be about 20 years higher than for a child born in 1929. The statistics in Table 3.1 are based on the death rates observed in 2019 for people of different ages. If people continue to live longer, Table 3.1 understates life expectancies.

Mortality risk is the risk that wars, epidemics such as AIDS, or pandemics such as COVID-19, which started at the end of 2019, will lead to people living not as long as expected. This adversely affects the payouts on most types of life insurance contracts (because the insured amount has to be paid earlier than expected), but should increase the profitability of annuity contracts (because the annuity is not paid out for as long). In calculating the impact of mortality risk, it is important to consider the age groups within the population that are likely to be most affected by a particular event.

To some extent, the longevity and mortality risks in the annuity business of a life insurance company offset those in its regular life insurance contracts. Actuaries must carefully assess the insurance company's net exposure under different scenarios. If the exposure is unacceptable, they may decide to enter into reinsurance contracts for some of the risks. Reinsurance is discussed later in this chapter.

3.4.1 Longevity Derivatives

A longevity derivative provides payoffs that are potentially attractive to insurance companies when they are concerned about their longevity exposure on annuity contracts and to pension funds. A typical contract is a *longevity bond*, also known as a *survivor bond*, which first traded in the late 1990s. A population group is defined and the coupon on the bond at any given time is defined as being proportional to the number of individuals in the population who are still alive.

Who will sell such bonds to insurance companies and pension funds? The answer is some speculators find the bonds attractive because they have very little systematic risk. (See Section 1.3 for a discussion of systematic risk.) The bond payments depend on how long people live and this is largely uncorrelated with returns from the market.

3.5 Property-Casualty Insurance

Property-casualty insurance can be subdivided into property insurance and casualty insurance. Property insurance provides protection against loss of or damage to property (from fire, theft, water damage, etc.). Casualty insurance provides protection against legal liability exposures (from, for example, injuries caused to third parties). Casualty insurance might more accurately be referred to as liability insurance. Sometimes both types of insurance are included in a single policy. For example, a home owner might buy insurance that provides protection against various types of loss such as property damage and theft as well as legal liabilities if others are injured while on the property. Similarly, car insurance typically provides protection against theft of, or damage to, one's own vehicle as well as protection against claims brought by others.

Typically, property-casualty policies are renewed from year to year and the insurance company will change the premium if its assessment of the expected payout changes. (This is different from life insurance, where premiums tend to remain the same for the life of the policy.) Because property-casualty insurance companies get involved in many different types of insurance, there is some natural risk diversification. Also, for some risks, the "law of large numbers" applies. For example, if an insurance company has written policies protecting 250,000 home owners against losses from theft and fire damage, the expected payout can be predicted reasonably accurately. This is because the policies provide protection against a large number of (almost) independent events. (Of course, there are liable to be trends through time in the number of losses and size of losses, and

the insurance company should keep track of these trends in determining year-to-year changes in the premiums.)

Property damage arising from natural disasters such as hurricanes, tornadoes, and earthquakes give rise to payouts for an insurance company that are much less easy to predict. These are termed *catastrophic risks*. The problem with them is that the claims made by different policyholders are not independent. Either a hurricane happens in a year and the insurance company has to deal with a large number of claims for hurricane-related damage or there is no hurricane in the year and therefore no claims are made. Most large insurers have models based on geographical, seismographical, and meteorological information to estimate the probabilities of catastrophes and the losses resulting therefrom. This provides a basis for setting premiums, but it does not alter the “all-or-nothing” nature of these risks for insurance companies.

Liability insurance, like catastrophe insurance, gives rise to total payouts that vary from year to year and are difficult to predict. For example, claims arising from asbestos-related damage to workers’ health have proved very expensive for insurance companies in the United States. A feature of liability insurance is what is known as *long-tail risk*. This is the possibility of claims being made several years after the insured period is over. In the case of asbestos, for example, the health risks were not realized until some time after exposure. As a result, the claims, when they were made, were under policies that had been in force several years previously. This creates a complication for actuaries and accountants. They cannot close the books soon after the end of each year and calculate a profit or loss. They must allow for the cost of claims that have not yet been made, but may be made some time in the future.

3.5.1 CAT Bonds

The derivatives market has come up with a number of products for hedging catastrophic risk. The most popular is a catastrophe (CAT) bond. This is a bond issued by a subsidiary of an insurance company that pays a higher-than-normal interest rate. In exchange for the extra interest, the holder of the bond agrees to cover payouts on a particular type of catastrophic risk that are in a certain range. Depending on the terms of the CAT bond, the interest or principal (or both) can be used to meet claims.

Suppose an insurance company has a \$70 million exposure to California earthquake losses and wants protection for losses over \$40 million. The insurance company could issue CAT bonds with a total principal of \$30 million. In the event that the insurance company’s California earthquake losses exceeded \$40 million, bondholders would lose some or all of their principal. As an alternative, the insurance company could cover the same losses by making a much bigger bond issue where only the bondholders’ interest is at risk. Yet another alternative is to make three separate bond issues covering losses in the range of \$40 to \$50 million, \$50 to \$60 million, and \$60 to \$70 million, respectively.

CAT bonds typically give a high probability of an above-normal rate of interest and a low probability of a high loss. Why would investors be interested in such instruments? The answer is that the return on CAT bonds, like the longevity bonds considered earlier, has no statistically significant correlations with market returns.¹ CAT bonds are therefore an attractive addition to an investor's portfolio. Their total risk can be completely diversified away in a large portfolio. If a CAT bond's expected return is greater than the risk-free interest rate (and typically it is), it has the potential to improve risk-return trade-offs.

3.5.2 Ratios Calculated by Property-Casualty Insurers

Insurance companies calculate a *loss ratio* for different types of insurance. This is the ratio of payouts made to premiums earned in a year. Loss ratios are typically in the 60% to 80% range. Statistics published by A. M. Best show that loss ratios in the United States have tended to increase through time. The *expense ratio* for an insurance company is the ratio of expenses to premiums earned in a year. The two major sources of expenses are loss adjustment expenses and selling expenses. Loss adjustment expenses are those expenses related to determining the validity of a claim and how much the policyholder should be paid. Selling expenses include the commissions paid to brokers and other expenses concerned with the acquisition of business. Expense ratios in the United States are typically in the 25% to 30% range and have tended to decrease through time.

The *combined ratio* is the sum of the loss ratio and the expense ratio. Suppose that for a particular category of policies in a particular year the loss ratio is 75% and the expense ratio is 30%. The combined ratio is then 105%. Sometimes a small dividend is paid to policyholders. Suppose that this is 1% of premiums. When this is taken into account we obtain what is referred to as the *combined ratio after dividends*. This is 106% in our example. This number suggests that the insurance company has lost 6% before tax on the policies being considered. In fact, this may not be the case. Premiums are generally paid by policyholders at the beginning of a year and payouts on claims are made during the year, or after the end of the year. The insurance company is therefore able to earn interest on the premiums during the time that elapses between the receipt of premiums and payouts. Suppose that, in our example, investment income is 9% of premiums received. When the investment income is taken into account, a ratio of $106 - 9 = 97\%$ is obtained. This is referred to as the *operating ratio*. Table 3.2 summarizes this example.

¹See R. H. Litzenberger, D. R. Beaglehole, and C. E. Reynolds, "Assessing Catastrophe Reinsurance-Linked Securities as a New Asset Class," *Journal of Portfolio Management* (Winter 1996): 76–86.

Table 3.2 Example Showing Calculation of Operating Ratio for a Property-Casualty Insurance Company

Loss ratio	75%
Expense ratio	<u>30%</u>
Combined ratio	105%
Dividends	<u>1%</u>
Combined ratio after dividends	106%
Investment income	<u>(9%)</u>
Operating ratio	97%

3.6 Health Insurance

Health insurance has some of the attributes of property-casualty insurance and some of the attributes of life insurance. It is sometimes considered to be a totally separate category of insurance. The extent to which health care is provided by the government varies from country to country. In the United States publicly funded health care has traditionally been limited and health insurance has therefore been an important consideration for most people. Canada is at the other extreme: nearly all health care needs are provided by a publicly funded system. Doctors are not allowed to offer most services privately. The main role of health insurance in Canada is to cover prescription costs and dental care, which are not funded publicly. In most other countries, there is a mixture of public and private health care. The United Kingdom, for example, has a publicly funded health care system, but some individuals buy insurance to have access to a private system that operates side by side with the public system. (The main advantage of private health insurance is a reduction in waiting times for routine elective surgery.)

In 2010, President Obama signed into law the Patient Protection and Affordable Care Act in an attempt to reform health care in the United States and increase the number of people with medical coverage. The eligibility for Medicaid (a program for low-income individuals) was expanded and subsidies were provided for low- and middle-income families to help them buy insurance. The act prevents health insurers from taking pre-existing medical conditions into account and requires employers to provide coverage to their employees or pay additional taxes. One difference between the United States and many other countries continues to be that health insurance is largely provided by the private rather than the public sector.

In health insurance, as in other forms of insurance, the policyholder makes regular premium payments and payouts are triggered by events. Examples of such events are the policyholder needing an examination by a doctor, the policyholder requiring treatment at a hospital, and the policyholder requiring prescription medication. Typically the premiums increase because of overall increases in the costs of providing health care. However, they usually cannot increase because the health of the policyholder deteriorates. It is interesting to compare health insurance with auto insurance and life insurance in this

respect. An auto insurance premium can increase (and usually does) if the policyholder's driving record indicates that expected payouts have increased and if the costs of repairs to automobiles have increased. Life insurance premiums do not increase—even if the policyholder is diagnosed with a health problem that significantly reduces life expectancy. Health insurance premiums are like life insurance premiums in that changes to the insurance company's assessment of the risk of a payout do not lead to an increase in premiums. However, it is like auto insurance in that increases in the overall costs of meeting claims do lead to premium increases.

Of course, when a policy is first issued, an insurance company does its best to determine the risks it is taking on. In the case of life insurance, questions concerning the policyholder's health have to be answered, pre-existing medical conditions have to be declared, and physical examinations may be required. In the case of auto insurance, the policyholder's driving record is investigated. In both of these cases, insurance can be refused. In the case of health insurance, legislation sometimes determines the circumstances under which insurance can be refused. As indicated earlier, the Patient Protection and Affordable Health Care Act makes it very difficult for insurance companies in the United States to refuse coverage because of pre-existing medical conditions.

Health insurance is often provided by the *group health insurance plans* of employers. These plans typically cover the employee and the employee's family. The cost of the health insurance is sometimes split between the employer and employee. The expenses that are covered vary from plan to plan. In the United States, most plans cover basic medical needs such as medical check-ups, physicals, treatments for common disorders, surgery, and hospital stays. Pregnancy costs may or may not be covered. Procedures such as cosmetic surgery are usually not covered.

3.7 Moral Hazard and Adverse Selection

We now consider two key risks facing insurance companies: moral hazard and adverse selection.

3.7.1 Moral Hazard

Moral hazard is the risk that the existence of insurance will cause the policyholder to behave differently than he or she would without the insurance. This different behavior increases the risks and the expected payouts of the insurance company. Three examples of moral hazard are:

1. A car owner buys insurance to protect against the car being stolen. As a result of the insurance, he or she becomes less likely to lock the car.
2. An individual purchases health insurance. As a result of the existence of the policy, more health care is demanded than previously.
3. As a result of a government-sponsored deposit insurance plan, a bank takes more risks because it knows that it is less likely to lose depositors because of this strategy. (This was discussed in Section 2.3.)

Moral hazard is not a big problem in life insurance. Insurance companies have traditionally dealt with moral hazard in property-casualty and health insurance in a number of ways. Typically there is a *deductible*. This means that the policyholder is responsible for bearing the first part of any loss. Sometimes there is a *co-insurance provision* in a policy. The insurance company then pays a predetermined percentage (less than 100%) of losses in excess of the deductible. In addition there is nearly always a *policy limit* (i.e., an upper limit to the payout). The effect of these provisions is to align the interests of the policyholder more closely with those of the insurance company.

3.7.2 *Adverse Selection*

Adverse selection is the phrase used to describe the problems an insurance company has when it cannot distinguish between good and bad risks. It offers the same price to everyone and inadvertently attracts more of the bad risks. If an insurance company is not able to distinguish good drivers from bad drivers and offers the same auto insurance premium to both, it is likely to attract more bad drivers. If it is not able to distinguish healthy from unhealthy people and offers the same life insurance premiums to both, it is likely to attract more unhealthy people.

To lessen the impact of adverse selection, an insurance company tries to find out as much as possible about the policyholder before committing itself. Before offering life insurance, it often requires the policyholder to undergo a physical examination by an approved doctor. Before offering auto insurance to an individual, it will try to obtain as much information as possible about the individual's driving record. In the case of auto insurance, it will continue to collect information on the driver's risk (number of accidents, number of speeding tickets, etc.) and make year-to-year changes to the premium to reflect this.

Adverse selection can never be completely overcome. It is interesting that, in spite of the physical examinations that are required, individuals buying life insurance tend to die earlier than mortality tables would suggest. But individuals who purchase annuities tend to live longer than mortality tables would suggest.

3.8 Reinsurance

Reinsurance is an important way in which an insurance company can protect itself against large losses by entering into contracts with another insurance company. For a fee, the second insurance company agrees to be responsible for some of the risks that have been insured by the first company. Reinsurance allows insurance companies to write more policies than they would otherwise be able to. Some of the counterparties in reinsurance contracts are other insurance companies or rich private individuals; others are companies that specialize in reinsurance such as Swiss Re and Warren Buffett's company, Berkshire Hathaway.

Reinsurance contracts can take a number of forms. Suppose that an insurance company has an exposure of \$100 million to hurricanes in Florida and wants to limit this to \$50 million. One alternative is to enter into annual reinsurance contracts that cover on a pro rata basis 50% of its exposure. (The reinsurer would then probably receive 50% of the premiums.) If hurricane claims in a particular year total \$70 million, the costs to the insurance company would be only $0.5 \times \$70$ or \$35 million, and the reinsurance company would pay the other \$35 million.

Another, more popular alternative, involving lower reinsurance premiums, is to buy a series of reinsurance contracts covering what are known as *excess cost layers*. The first layer might provide indemnification for losses between \$50 million and \$60 million, the next layer might cover losses between \$60 million and \$70 million, and so on. Each reinsurance contract is known as an *excess-of-loss* reinsurance contract.

3.9 Capital Requirements

The balance sheets for life insurance and property-casualty insurance companies are different because the risks taken and the reserves that must be set aside for future payouts are different.

3.9.1 Life Insurance Companies

Table 3.3 shows an abbreviated balance sheet for a life insurance company. Most of the life insurance company's investments are in corporate bonds. The insurance company tries to match the maturity of its assets with the maturity of liabilities. However, it takes on credit risk because the default rate on the bonds may be higher than expected.

Unlike a bank, an insurance company has exposure on the liability side of the balance sheet as well as on the asset side. The policy reserves (80% of assets in this case) are estimates (usually conservative) of actuaries for the present value of payouts on the policies that have been written. The estimates may prove to be low if the holders of life insurance policies die earlier than expected or the holders of annuity contracts live longer than expected. The 10% equity on the balance sheet includes the original equity contributed and retained earnings and provides a cushion. If payouts are greater than loss reserves by an amount equal to 5% of assets, equity will decline, but the life insurance company will survive.

Table 3.3 Abbreviated Balance Sheet for Life Insurance Company

Assets		Liabilities and Net Worth	
Investments	90	Policy reserves	80
Other assets	10	Subordinated long-term debt	10
		Equity capital	10
Total	<u>100</u>	Total	<u>100</u>

Table 3.4 Abbreviated Balance Sheet for Property-Casualty Insurance Company

Assets		Liabilities and Net Worth	
Investments	90	Policy reserves	45
Other assets	10	Unearned premiums	15
		Subordinated long-term debt	10
		Equity capital	30
Total	100	Total	100

3.9.2 *Property-Casualty Insurance Companies*

Table 3.4 shows an abbreviated balance sheet for a property-casualty life insurance company. A key difference between Table 3.3 and Table 3.4 is that the equity in Table 3.4 is much higher. This reflects the differences in the risks taken by the two sorts of insurance companies. The payouts for a property-casualty company are much less easy to predict than those for a life insurance company. Who knows when a hurricane will hit Miami or how large payouts will be for the next asbestos-like liability problem? The unearned premiums item on the liability side represents premiums that have been received, but apply to future time periods. If a policyholder pays \$2,500 for house insurance on June 30 of a year, only \$1,250 has been earned by December 31 of the year. The investments in Table 3.4 consist largely of liquid bonds with shorter maturities than the bonds in Table 3.3.

3.10 The Risks Facing Insurance Companies

The most obvious risk for an insurance company is that the policy reserves are not sufficient to meet the claims of policyholders. Although the calculations of actuaries are usually fairly conservative, there is always the chance that payouts much higher than anticipated will be required. Insurance companies also face risks concerned with the performance of their investments. Many of these investments are in corporate bonds. If defaults on corporate bonds are above average, the profitability of the insurance company will suffer. It is important that an insurance company's bond portfolio be diversified by business sector and geographical region. An insurance company also needs to monitor the liquidity risks associated with its investments. Illiquid bonds (e.g., those the insurance company might buy in a private placement) tend to provide higher yields than bonds that are publicly owned and actively traded. However, they cannot be as readily converted into cash to meet unexpectedly high claims. Insurance companies enter into transactions with banks and reinsurance companies. This exposes them to credit risk. Like banks, insurance companies are also exposed to operational risks and business risks.

Regulators specify minimum capital requirements for an insurance company to provide a cushion against losses. Insurance companies, like banks, have also developed their own procedures for calculating economic capital. This is their own internal estimate of required capital (see Chapter 28).

3.11 Regulation

The ways in which insurance companies are regulated in the United States and Europe are quite different.

3.11.1 *United States*

In the United States, the McCarran-Ferguson Act of 1945 confirmed that insurance companies are regulated at the state level rather than the federal level. (Banks, by contrast, are regulated at the federal level.) State regulators are concerned with the solvency of insurance companies and their ability to satisfy policyholders' claims. They are also concerned with business conduct (i.e., how premiums are set, advertising, contract terms, the licensing of insurance agents and brokers, and so on).

The National Association of Insurance Commissioners (NAIC) is an organization consisting of the chief insurance regulatory officials from all 50 states. It provides a national forum for insurance regulators to discuss common issues and interests. It also provides some services to state regulatory commissions. For example, it provides statistics on the loss ratios of property-casualty insurers. This helps state regulators identify those insurers for which the ratios are outside normal ranges.

Insurance companies are required to file detailed annual financial statements with state regulators, and the state regulators conduct periodic on-site reviews. Capital requirements are determined by regulators using risk-based capital standards determined by NAIC. These capital levels reflect the risk that policy reserves are inadequate, that counterparties in transactions default, and that the return on investments are less than expected.

The policyholder is protected against an insurance company becoming insolvent (and therefore unable to make payouts on claims) by insurance guaranty associations. An insurer is required to be a member of the guaranty association in a state as a condition of being licensed to conduct business in the state. When there is an insolvency by another insurance company operating in the state, each insurance company operating in the state has to contribute an amount to the state guaranty fund that is dependent on the premium income it collects in the state. The fund is used to pay the small policyholders of the insolvent insurance company. (The definition of a small policyholder varies from state to state.) There may be a cap on the amount the insurance company has to contribute to the state guaranty fund in a year. This can lead to the policyholder having to wait several years before the guaranty fund is in a position to make a full payout on its claims. In the case of life insurance, where policies last for many years, the policyholders of insolvent companies are usually taken over by other insurance companies. However, there may be some change to the terms of the policy so that the policyholder is somewhat worse off than before.

The guaranty system for insurance companies in the United States is therefore different from that for banks. In the case of banks, there is a permanent fund created from premiums paid by banks to the FDIC to protect depositors. In the case of insurance companies, there is no permanent fund. Insurance companies have to make contributions

after an insolvency has occurred. An exception to this is property-casualty companies in New York State, where a permanent fund does exist.

Regulating insurance companies at the state level is unsatisfactory in some respects. Regulations are not uniform across the different states. A large insurance company that operates throughout the United States has to deal with a large number of different regulatory authorities. Some insurance companies trade derivatives in the same way as banks, but are not subject to the same regulations as banks. This can create problems. In 2008, it transpired that a large insurance company, American International Group (AIG), incurred huge losses trading credit derivatives and had to be bailed out by the federal government.

The Dodd–Frank Act of 2010 set up the Federal Insurance Office (FIO), which is housed in the Department of the Treasury. It is tasked with monitoring the insurance industry and identifying gaps in regulation. It can recommend to the Financial Stability Oversight Council that a large insurance company (such as AIG) be designated as a nonbank financial company supervised by the Federal Reserve. It also liaises with regulators in other parts of the world (particularly, those in the European Union) to foster the convergence of regulatory standards. It may be the case that the United States will eventually move to a federal system for insurance regulation.

3.11.2 Europe

In the European Union, insurance companies are regulated centrally. This means that in theory the same regulatory framework applies to insurance companies throughout all member countries. The original framework (established in the 1970s) is known as Solvency I. It was heavily influenced by research carried out by Professor Cornelis Campagne from the Netherlands who showed that, with a capital equal to 4% of policy provisions, life insurance companies have a 95% chance of surviving. Investment risks are not explicitly considered by Solvency I.

In 2016, the European Union replaced Solvency I by Solvency II. This assigns capital for a wider range of risks than Solvency I and will be discussed in Chapter 25.

3.12 Pension Plans

Pension plans are set up by companies for their employees. Typically, contributions are made to a pension plan by both the employee and the employer while the employee is working. When the employee retires, he or she receives a pension until death. A pension fund therefore involves the creation of a lifetime annuity from regular contributions and has similarities to some of the products offered by life insurance companies. There are two types of pension plans: defined benefit and defined contribution.

In a *defined benefit plan*, the pension that the employee will receive on retirement is defined by the plan. Typically it is calculated by a formula that is based on the number of years of employment and the employee's salary. For example, the pension per year might

equal the employee's average earnings per year during the last three years of employment multiplied by the number of years of employment multiplied by 2%. The employee's spouse may continue to receive a (usually reduced) pension if the employee dies before the spouse. In the event of the employee's death while still employed, a lump sum is often payable to dependents and a monthly income may be payable to a spouse or dependent children. Sometimes pensions are adjusted for inflation. This is known as indexation. For example, the indexation in a defined benefit plan might lead to pensions being increased each year by 75% of the increase in the consumer price index. Pension plans that are sponsored by governments (such as Social Security in the United States) are similar to defined benefit plans in that they require regular contributions up to a certain age and then provide lifetime pensions.

In a *defined contribution plan* the employer and employee contributions are invested on behalf of the employee. When employees retire, there are typically a number of options open to them. The amount to which the contributions have grown can be converted to a lifetime annuity. In some cases, the employee can opt to receive a lump sum instead of an annuity.

The key difference between a defined contribution and a defined benefit plan is that, in the former, the funds are identified with individual employees. An account is set up for each employee and the pension is calculated only from the funds contributed to that account. By contrast, in a defined benefit plan, all contributions are pooled and payments to retirees are made out of the pool. In the United States, a 401(k) plan is a form of defined contribution plan where the employee elects to have some portion of his or her income directed to the plan (with possibly some employer matching) and can choose between a number of investment alternatives (e.g., stocks, bonds, and money market instruments).

An important aspect of both defined benefit and defined contribution plans is the deferral of taxes. No taxes are payable on money contributed to the plan by the employee and contributions by a company are deductible. Taxes are payable only when pension income is received (and at this time the employee may have a relatively low marginal tax rate).

Defined contribution plans involve very little risk for employers. If the performance of the plan's investments is less than anticipated, the employee bears the cost. By contrast, defined benefit plans impose significant risks on employers because they are ultimately responsible for paying the promised benefits. Let us suppose that the assets of a defined benefit plan total \$100 million and that actuaries calculate the present value of the obligations to be \$120 million. The plan is \$20 million underfunded and the employer is required to make up the shortfall (usually over a number of years). The risks posed by defined benefit plans have led some companies to convert defined benefit plans to defined contribution plans.

Estimating the present value of the liabilities in defined benefit plans is not easy. An important issue is the discount rate used. The higher the discount rate, the lower the present value of the pension plan liabilities. It used to be common to use the average rate of return on the assets of the pension plan as the discount rate. This encourages the

BUSINESS SNAPSHOT 3.2

A Perfect Storm

During the period from December 31, 1999, to December 31, 2002, the S&P 500 declined by about 40% from 1469.25 to 879.82, and the 20-year Treasury rate in the United States declined by 200 basis points from 6.83% to 4.83%. The impact of the first of these events was that the market value of the assets of defined benefit pension plans declined sharply. The impact of the second of the two events was that the discount rate used by defined benefit plans for their liabilities decreased so that the fair value of the liabilities calculated by actuaries increased. This created a “perfect storm” for the pension plans. Many funds that had been overfunded became underfunded. Funds that had been slightly underfunded became much more seriously underfunded.

When a company has a defined benefit plan, the value of its equity is adjusted to reflect the amount by which the plan is overfunded or underfunded. It is not surprising that many companies have tried to replace defined benefit pension plans with defined contribution plans to avoid the risk of equity being eroded by a perfect storm.

pension plan to invest in equities because the average return on equities is higher than the average return on bonds, making the value of the liabilities look low. Accounting standards now recognize that the liabilities of pension plans are obligations similar to bonds and require the liabilities of the pension plans of private companies to be discounted at AA-rated bond yields. The difference between the value of the assets of a defined benefit plan and that of its liabilities must be recorded as an asset or liability on the balance sheet of the company. Thus, if a company's defined benefit plan is underfunded, the company's shareholder equity is reduced. A perfect storm is created when the assets of a defined benefit pension plan decline sharply in value and the discount rate for its liabilities decreases sharply (see Business Snapshot 3.2).

3.12.1 *Are Defined Benefit Plans Viable?*

A typical defined benefit plan provides the employee with about 70% of final salary as a pension and includes some indexation for inflation. What percentage of the employee's income during his or her working life should be set aside for providing the pension? The answer depends on assumptions about interest rates, how fast the employee's income rises during the employee's working life, and so on. But, if an insurance company were asked to provide a quote for the sort of defined benefit plan we are considering, the required contribution rate would be about 25% of income each year. (Problems 3.15 and 3.19 provide an indication of calculations that can be carried out.) The insurance company

would invest the premiums in corporate bonds (in the same way that it does the premiums for life insurance and annuity contracts) because this provides the best way of matching the investment income with the payouts.

The contributions to defined benefit plans (employer plus employee) are much less than 25% of income. In a typical defined benefit plan, the employer and employee each contribute around 5%. The total contribution is therefore only 40% of what an insurance actuary would calculate the required premium to be. It is therefore not surprising that many defined pension plans are underfunded.

Unlike insurance companies, pension funds choose to invest a significant proportion of their assets in equities. (A typical portfolio mix for a pension plan is 60% equity and 40% debt.) By investing in equities, the pension fund is creating a situation where there is some chance that the pension plan will be fully funded. But there is also some chance of severe underfunding. If equity markets do well, as they have done from 1960 to 2020 in many parts of the world, defined benefit plans may find they can afford their liabilities. But if equity markets perform badly, there are likely to be problems.

This raises an interesting question: Who is responsible for underfunding in defined benefit plans? In the first instance, it is the company's shareholders who bear the cost. If the company declares bankruptcy, the cost may be borne by the government via insurance that is offered.² In either case there is a transfer of wealth to retirees from the next generation.

Many people argue that wealth transfers from one generation to another are not acceptable. A 25% contribution rate to pension plans is probably not feasible. If defined benefit plans are to continue, there must be modifications to the terms of the plans so that there is some risk sharing between retirees and the next generation. If equity markets perform badly during their working life, retirees must be prepared to accept a lower pension and receive only modest help from the next generation. If equity markets perform well, retirees can receive a full pension and some of the benefits can be passed on to the next generation.

Longevity risk is a major concern for pension plans. We mentioned earlier that life expectancy increased by about 20 years between 1929 and 2019. If this trend continues, the underfunding problems of defined benefit plans (both those administered by companies and those administered by national governments) may become more severe. It is not surprising that, in many jurisdictions, individuals have the right to work past the normal retirement age. Depending on the rules governing pensions in a particular jurisdiction, this can help solve the problems faced by defined benefit pension plans. An individual who retires at 70 rather than 65 makes an extra five years of pension contributions and the period of time for which the pension is received is shorter by five years.

²For example, in the United States, the Pension Benefit Guaranty Corporation (PBGC) insures private defined benefit plans. If the premiums the PBGC receives from plans are not sufficient to meet claims, presumably the government would have to step in.

Summary

There are two main types of insurance companies: life and property-casualty. Life insurance companies offer a number of products that provide a payoff when the policyholder dies. Term life insurance provides a payoff only if the policyholder dies during a certain period. Whole life insurance provides a payoff on the death of the insured, regardless of when this is. There is a savings element to whole life insurance. Typically, the portion of the premium not required to meet expected payouts in the early years of the policy is invested, and this is used to finance expected payouts in later years. Whole life insurance policies usually give rise to tax benefits, because the present value of the tax paid is less than it would be if the investor had chosen to invest funds directly rather than through the insurance policy, and in some jurisdictions is zero.

Life insurance companies also offer annuity contracts. These are contracts that, in return for a lump-sum payment, provide the policyholder with an annual income from a certain date for the rest of his or her life. Mortality tables provide important information for the valuation of the life insurance contracts and annuities. However, actuaries must consider (a) longevity risk (the possibility that people will live longer than expected) and (b) mortality risk (the possibility that epidemics such as COVID-19 will reduce life expectancy for some segments of the population).

Property-casualty insurance is concerned with providing protection against a loss of, or damage to, property. It also protects individuals and companies from legal liabilities. The most difficult payouts to predict are those where the same event is liable to trigger claims by many policyholders at about the same time. Examples of such events are hurricanes or earthquakes.

Health insurance has some of the features of life insurance and some of the features of property-casualty insurance. Health insurance premiums are like life insurance premiums in that changes to the company's assessment of the risk of payouts do not lead to an increase in premiums. However, it is like property-casualty insurance in that increases in the overall costs of providing health care can lead to increases in premiums.

Two key risks in insurance are moral hazard and adverse selection. Moral hazard is the risk that the behavior of an individual or corporation with an insurance contract will be different from the behavior without the insurance contract. Adverse selection is the risk that the individuals and companies who buy a certain type of policy are those for which expected payouts are relatively high. Insurance companies take steps to reduce these two types of risk, but they cannot eliminate them altogether.

Insurance companies are different from banks in that their liabilities as well as their assets are subject to risk. A property-casualty insurance company must typically keep more equity capital, as a percent of total assets, than a life insurance company. In the United States, insurance companies are different from banks in that they are regulated at the state level rather than at the federal level. In Europe, insurance companies are regulated by the European Union and by national governments. The European Union has developed a new set of capital requirements, known as Solvency II.

There are two types of pension plans: defined benefit plans and defined contribution plans. Defined contribution plans are straightforward. Contributions made by an employee and contributions made by the company on behalf of the employee are kept in a separate account, invested on behalf of the employee, and converted into a lifetime annuity when the employee retires. In a defined benefit plan, contributions from all employees and the company are pooled and invested. Retirees receive a pension that is based on the salary they earned while working. The viability of defined benefit plans is questionable. Many are underfunded and need superior returns from equity markets to pay promised pensions to both current and future retirees.

Further Reading

- Ambachtsheer, K. P. *Pension Revolution: A Solution to the Pensions Crisis*. Hoboken, NJ: John Wiley & Sons, 2007.
- Canter, M. S., J. B. Cole, and R. L. Sandor. "Insurance Derivatives: A New Asset Class for the Capital Markets and a New Hedging Tool for the Insurance Industry." *Journal of Applied Corporate Finance* (Autumn 1997): 69–83.
- Doff, R. *Risk Management for Insurers: Risk Control, Economic Capital, and Solvency II*. London: Risk Books, 2007.
- Federal Insurance Office. "How to Modernize and Improve the System of Insurance Regulation in the United States." Report, December 2013.
- Froot, K. A. "The Market for Catastrophe Risk: A Clinical Examination." *Journal of Financial Economics* 60 (2001): 529–571.
- Litzenberger, R. H., D. R. Beaglehole, and C. E. Reynolds. "Assessing Catastrophe Reinsurance-Linked Securities as a New Asset Class." *Journal of Portfolio Management* (Winter 1996): 76–86.

Practice Questions and Problems (Answers at End of Book)

- 3.1 What is the difference between term life insurance and whole life insurance?
- 3.2 Explain the meaning of variable life insurance and universal life insurance.
- 3.3 A life insurance company offers whole life and annuity contracts. In which contracts does it have exposure to (a) longevity risk, (b) mortality risk?
- 3.4 "Equitable Life gave its policyholders a free option." Explain the nature of the option.
- 3.5 Use Table 3.1 to calculate the minimum premium an insurance company should charge for a \$1 million two-year term life insurance policy issued to a woman age 50. Assume that the premium is paid at the beginning of each year and that the interest rate is zero.
- 3.6 From Table 3.1, what is the probability that a man age 30 will live to 90? What is the same probability for a woman age 30?
- 3.7 What features of the policies written by a property-casualty insurance company give rise to the most risk?

- 3.8 Explain how CAT bonds work.
- 3.9 Consider two bonds that have the same coupon, time to maturity, and price. One is a B-rated corporate bond. The other is a CAT bond. An analysis based on historical data shows that the expected losses on the two bonds in each year of their life is the same. Which bond would you advise a portfolio manager to buy, and why?
- 3.10 How does health insurance in the United States differ from that in Canada and the United Kingdom?
- 3.11 An insurance company decides to offer individuals insurance against losing their jobs. What problems is it likely to encounter?
- 3.12 Why do property-casualty insurance companies hold more capital than life insurance companies?
- 3.13 Explain what is meant by “loss ratio” and “expense ratio” for a property-casualty insurance company. “If an insurance company is profitable, it must be the case that the loss ratio plus the expense ratio is less than 100%.” Discuss this statement.
- 3.14 What is the difference between a defined benefit and a defined contribution pension plan?
- 3.15 Suppose that in a certain defined benefit pension plan
- (a) Employees work for 40 years earning wages that increase with inflation.
 - (b) They retire with a pension equal to 75% of their final salary. This pension also increases with inflation.
 - (c) The pension is received for 20 years.
 - (d) The pension fund’s income is invested in bonds that earn the inflation rate.
- Estimate the percentage of an employee’s salary that must be contributed to the pension plan if it is to remain solvent. (*Hint: Do all calculations in real rather than nominal dollars.*)

Further Questions

- 3.16 Use Table 3.1 to calculate the minimum premium an insurance company should charge for a \$5 million three-year term life insurance contract issued to a man age 60. Assume that the premium is paid at the beginning of each year and death always takes place halfway through a year. The risk-free interest rate is 6% per annum (with semiannual compounding).
- 3.17 An insurance company’s losses of a particular type per year are to a reasonable approximation normally distributed with a mean of \$150 million and a standard deviation of \$50 million. (Assume that the risks taken by the insurance company are entirely nonsystematic.) The one-year risk-free rate is 5% per annum with annual compounding. Estimate the cost of the following:
- (a) A contract that will pay in one-year’s time 60% of the insurance company’s costs on a pro rata basis.
 - (b) A contract that pays \$100 million in one-year’s time if losses exceed \$200 million.

- 3.18 During a certain year, interest rates fall by 200 basis points (2%) and equity prices are flat. Discuss the effect of this on a defined benefit pension plan that is 60% invested in equities and 40% invested in bonds.
- 3.19 Suppose that in a certain defined benefit pension plan
- (a) Employees work for 45 years earning wages that increase at a real rate of 2%.
 - (b) They retire with a pension equal to 70% of their final salary. This pension increases at the rate of inflation minus 1%.
 - (c) The pension is received for 18 years.
 - (d) The pension fund's income is invested in bonds that earn the inflation rate plus 1.5%.

Estimate the percentage of an employee's salary that must be contributed to the pension plan if it is to remain solvent. (*Hint: Do all calculations in real rather than nominal dollars.*)

Chapter 4

Fund Managers

Fund managers invest money on behalf of individuals and companies. The funds from different investors are pooled and investments are chosen by the fund manager in an attempt to meet specified objectives. Mutual funds (which are called “unit trusts” in some countries) and ETFs serve the needs of relatively small investors, while hedge funds seek to attract funds from wealthy individuals and large investors such as pension funds. Hedge funds are subject to much less regulation than mutual funds and ETFs. They are free to use a wider range of trading strategies than mutual funds and are usually more secretive about what they do. Mutual funds and ETFs are required to explain their investment policies in a prospectus that is available to potential investors.

This chapter describes the types of mutual funds, ETFs, and hedge funds that exist. It examines how they are regulated and the fees they charge. It also looks at how successful they have been at producing good returns for investors.

4.1 Mutual Funds

One of the attractions of mutual funds for the small investor is the diversification opportunities they offer. As we saw in Chapter 1, diversification improves an investor’s risk-return trade-off. However, it can be difficult for a small investor to hold enough stocks to be well diversified. In addition, maintaining a well-diversified portfolio can lead to high transaction costs. A mutual fund provides a way in which the resources of

Table 4.1 Growth of Assets of Open-End Mutual Funds in the United States

Year	Assets (\$ billions)
1940	0.5
1960	17.0
1980	134.8
2000	6,964.6
2021	26,389.9

SOURCE: Investment Company Institute.

many small investors are pooled so that the benefits of diversification are realized at a relatively low cost.

An investor in a mutual fund owns a certain number of shares in the fund. The most common type of mutual fund is an *open-end fund*. This means that the total number of shares outstanding goes up as investors buy more shares and down as shares are redeemed. Open-end mutual funds have grown very fast since the Second World War. Table 4.1 shows estimates of the assets managed by open-end mutual funds in the United States since 1940. These assets had grown to over \$26 trillion by 2021. Over 45% of U.S. households own mutual funds. Some mutual funds are offered by firms that specialize in asset management, such as Fidelity. Others are offered by banks such as JPMorgan Chase. The largest fund manager in the world is BlackRock whose assets under management passed \$10 trillion at the end of 2021.

Money market mutual funds invest in interest-bearing instruments, such as Treasury bills, commercial paper, and bankers' acceptances, with a life of less than one year. They are an alternative to interest-bearing bank accounts and usually provide a higher rate of interest because they are not insured by a government agency. Some money market funds offer check-writing facilities similar to banks. Money market fund investors are typically risk-averse and do not expect to lose any of the funds invested. In other words, investors expect a positive return after management fees.¹ In normal market conditions this is what they get. But occasionally the return is negative so that some principal is lost. This is known as "breaking the buck" because a \$1 investment is then worth less than \$1. After Lehman Brothers defaulted in September 2008, the oldest money market fund in the United States, Reserve Primary Fund, broke the buck because it had to write off short-term debt issued by Lehman. To avoid a run on money market funds (which would have meant that healthy companies had no buyers for their commercial paper), a government-backed guaranty program was introduced. It lasted for about a year.

¹Stable value funds are a popular alternative to money market funds. They typically invest in bonds and similar instruments with lives of up to five years. Banks and other companies provide (for a price) insurance guaranteeing that the return will not be negative.

There are three main types of long-term funds:

1. Bond funds that invest in fixed-income securities with a life of more than one year.
2. Equity funds that invest in common and preferred stock.
3. Hybrid funds that invest in stocks, bonds, and other securities.

Equity mutual funds are by far the most popular.

Mutual funds are valued at 4 P.M. each day. This involves the mutual fund manager calculating the market value of each asset in the portfolio so that the total value of the fund is determined. This total value is divided by the number of shares outstanding to obtain the value of each share. The latter is referred to as the *net asset value* (NAV) of the fund. Shares in the fund can be bought from the fund or sold back to the fund. When an investor issues instructions to buy or sell shares, it is the next-calculated NAV that applies to the transaction. For example, if an investor decides to buy at 2 P.M. on a particular business day, the NAV at 4 P.M. on that day determines the amount paid by the investor.

4.1.1 Index Funds

Some funds are designed to track a particular equity index, such as the S&P 500 or the FTSE 100. The tracking can most simply be achieved by buying all the shares in the index in amounts that reflect their weight. For example, if IBM has 1% weight in a particular index, 1% of the tracking portfolio for the index would be invested in IBM stock. Another way of achieving tracking is to choose a smaller portfolio of representative shares that has been shown by research to track the chosen portfolio closely. Yet another way is to use index futures.

One of the first index funds was launched in the United States on December 31, 1975, by John Bogle to track the S&P 500. It started with only \$11 million of assets and was initially ridiculed as being “un-American” and “Bogle’s folly.” However, it has been hugely successful and has been renamed the Vanguard 500 Index Fund (symbol: VFINX). Its assets under management had grown to about \$280 billion by the end of 2021.

How accurately do index funds track the index? Two relevant measures are the *tracking error* and the *expense ratio*. The tracking error of a fund can be defined as either the root mean square error of the difference between the fund’s return per year and the index return per year or as the standard deviation of this difference.² The expense ratio is the fee charged per year, as a percentage of assets, for administering the fund.

4.1.2 Costs

Mutual funds incur a number of different costs. These include management expenses, sales commissions, accounting and other administrative costs, transaction costs on trades,

²The root mean square error of the difference (square root of the average of the squared differences) is a better measure. The trouble with standard deviation is that it is low when the errors are large, of the same sign, and fairly constant.

Table 4.2 Average Total Cost per Year When Mutual Fund Is Held for Five Years (% of assets)

Country	Bond Funds	Equity Funds
Australia	0.75	1.41
Austria	1.55	2.37
Belgium	1.60	2.27
Canada	1.84	3.00
Denmark	1.91	2.62
Finland	1.76	2.77
France	1.57	2.31
Germany	1.48	2.29
Italy	1.56	2.58
Luxembourg	1.62	2.43
Netherlands	1.73	2.46
Norway	1.77	2.67
Spain	1.58	2.70
Sweden	1.67	2.47
Switzerland	1.61	2.40
United Kingdom	1.73	2.48
United States	<u>1.05</u>	<u>1.53</u>
Average	1.39	2.09

SOURCE: A. Khorana, H. Servaes, and P. Tufano, "Mutual Fund Fees Around the World," *Review of Financial Studies* 22 (March 2009): 1279–1310.

and so on. To recoup these costs, and to make a profit, fees are charged to investors. A *front-end load* is a fee charged when an investor first buys shares in a mutual fund. Not all funds charge this type of fee. Those that do are referred to as front-end loaded. In some jurisdictions, front-end load fees are restricted to being less than a certain amount. Some funds charge fees when an investor sells shares. These are referred to as a *back-end load*. Typically the back-end load declines with the length of time the shares in the fund have been held. All funds charge an annual fee. There may be separate fees to cover management expenses, distribution costs, and so on. The *total expense ratio* is the total of the annual fees charged per share divided by the value of the share.

Mutual fund fees vary widely from country to country. This was illustrated by Khorana et al. (2009).³ They assumed that a fund is kept for five years. The total shareholder cost per year was calculated as

$$\text{Total expense ratio} + \frac{\text{Front-end load}}{5} + \frac{\text{Back-end load}}{5}$$

Their results are summarized in Table 4.2. A more recent analysis has been carried out by Morningstar.⁴ It concluded that Australia, the Netherlands, and the United States had

³See A. Khorana, H. Servaes, and P. Tufano, "Mutual Fund Fees Around the World," *Review of Financial Studies* 22 (March 2009): 1279–1310.

⁴See <https://www.morningstar.com/insights/2019/09/17/global-mutual-funds>.

the lowest fees, while Italy and Taiwan had the highest. Index funds tend to have lower fees than regular funds because no highly paid stock pickers or analysts are required. For some index funds in the United States, total fees are as low as 0.15% per year.

4.1.3 Closed-End Funds

The funds we have talked about so far are open-end funds. These are by far the most common type of fund. The number of shares outstanding varies from day to day as individuals choose to invest in the fund or redeem their shares. Closed-end funds are like regular corporations and have a fixed number of shares outstanding. The shares of the fund are traded on a stock exchange. For closed-end funds, two NAVs can be calculated. One is the price at which the shares of the fund are trading. The other is the market value of the fund's portfolio divided by the number of shares outstanding. The latter can be referred to as the fair market value. Usually a closed-end fund's share price is less than its fair market value. A number of researchers have investigated the reason for this. Research by Ross (2002) suggests that the fees paid to fund managers provide the explanation.⁵ Closed-end funds are much less popular than open-end funds. In 2021, their assets in the United States totaled about \$309 billion.

4.2 Exchange-Traded Funds

Exchange-traded funds (ETFs) have existed in the United States since 1993 and in Europe since 1999. They often track an index and so are an alternative to an index mutual fund for investors who are comfortable earning a return that is designed to mirror the index. One of the most widely known ETFs, called the Spider, tracks the S&P 500 and trades under the symbol SPY. In a survey of investment professionals conducted in March 2008, 67% called ETFs the most innovative investment vehicle of the previous two decades and 60% reported that ETFs have fundamentally changed the way they construct investment portfolios. In 2008, the SEC in the United States authorized the creation of actively managed ETFs.

ETFs are created by institutional investors. Typically, an institutional investor deposits a block of securities with the ETF and obtains shares in the ETF (known as *creation units*) in return. Some or all of the shares in the ETF are then traded on a stock exchange. This gives ETFs the characteristics of a closed-end fund rather than an open-end fund. However, a key feature of ETFs is that institutional investors can exchange large blocks of shares in the ETF for the assets underlying the shares at that time. They can give up

⁵See S. Ross, "Neoclassical Finance, Alternative Finance and the Closed End Fund Puzzle," *European Financial Management* 8 (2002): 129–137.

shares they hold in the ETF and receive the assets or they can deposit new assets and receive new shares. This ensures that there is never any appreciable difference between the price at which shares in the ETF are trading on the stock exchange and their fair market value. This is a key difference between ETFs and closed-end funds and makes ETFs more attractive to investors than closed-end funds.

ETFs have a number of advantages over open-end mutual funds. ETFs can be bought or sold at any time of the day. They can be shorted in the same way that shares in any stock are shorted. (See Chapter 5 for a discussion of short selling.) ETF holdings are disclosed twice a day, giving investors full knowledge of the assets underlying the fund. Mutual funds by contrast only have to disclose their holdings relatively infrequently. When shares in a mutual fund are sold, managers often have to sell the stocks in which the fund has invested to raise the cash that is paid to the investor. When shares in the ETF are sold, this is not necessary as another investor is providing the cash. This means that transaction costs are saved and there are less unplanned capital gains and losses passed on to shareholders. Finally, the expense ratios of ETFs tend to be lower than those of mutual funds. The popularity of ETFs has been increasing. By 2020, the assets of ETFs had grown to \$7.7 trillion globally.

4.3 Active vs. Passive Management

When buying a mutual fund or an ETF, an investor has a choice between a passively managed fund that is designed to track an index such as the S&P 500 and an actively managed fund that relies on the stock selection and timing skills of the fund manager. Actively managed funds tend to have much higher expense ratios. A key question therefore is whether actively managed mutual funds outperform stock indices such as the S&P 500. Some funds in some years do very well, but this could be the result of good luck rather than good investment management. Two key questions for researchers are:

1. Do actively managed funds outperform stock indices on average?
2. Do funds that outperform the market in one year continue to do so?

The answer to both questions appears to be no. In a classic study, Jensen (1969) performed tests on mutual fund performance using 10 years of data on 115 funds.⁶ He calculated the alpha for each fund in each year. (As explained in Section 1.3, alpha is the return earned in excess of that predicted by the capital asset pricing model.) The average alpha was about zero before all expenses and negative after expenses were considered. Jensen tested whether funds with positive alphas tended to continue to earn positive alphas. His results are summarized in Table 4.3. The first row shows that 574 positive

⁶See M. C. Jensen, "Risk, the Pricing of Capital Assets and the Evaluation of Investment Portfolios," *Journal of Business* 42 (April 1969): 167–247.

Table 4.3 Consistency of Good Performance by Mutual Funds

Number of Consecutive Years of Positive Alpha	Number of Observations	Percentage of Observations When Next Alpha Is Positive
1	574	50.4
2	312	52.0
3	161	53.4
4	79	55.8
5	41	46.4
6	17	35.3

alphas were observed from the 1,150 observations (close to 50%). Of these positive alphas, 50.4% were followed by another year of positive alpha. Row two shows that, when two years of positive alphas have been observed, there is a 52% chance that the next year will have a positive alpha, and so on. The results show that, when a manager has achieved above-average returns for one year (or several years in a row), there is still only a probability of about 50% of achieving above-average returns the next year. The results suggest that managers who obtain positive alphas do so because of luck rather than skill. It is possible that there are some managers who are able to perform consistently above average, but they are a very small percentage of the total.

More recent studies have confirmed Jensen's conclusions. On average, mutual fund managers do not beat the market and past performance is not a good guide to future performance. The success of index funds shows that this research has influenced the views of many investors.

Mutual funds frequently advertise impressive returns. However, the fund being featured might be one fund out of many offered by the same organization that happens to have produced returns well above the average for the market. Distinguishing between good luck and good performance is always tricky. Suppose an asset management company has 32 funds following different trading strategies and assume that the fund managers have no particular skills, so that the return of each fund has a 50% chance of being greater than the market each year. The probability of a particular fund beating the market every year for the next five years is $(1/2)^5$ or $1/32$. This means that by chance one out of the 32 funds will show a great performance over the five-year period!

One point should be made about the way returns over several years are expressed. One mutual fund might advertise "The average of the returns per year that we have achieved over the last five years is 15%." Another might say "If you had invested your money in our mutual fund for the last five years your money would have grown at 15% per year." These statements sound the same, but are actually different, as illustrated by Business Snapshot 4.1. In many countries, regulators have strict rules to ensure that mutual fund returns are not reported in a misleading way.

BUSINESS SNAPSHOT 4.1

Mutual Fund Returns Can Be Misleading

Suppose that the following is a sequence of returns per annum reported by a mutual fund manager over the last five years (measured using annual compounding):

$$15\%, \quad 20\%, \quad 30\%, \quad -20\%, \quad 25\%$$

The arithmetic mean of the returns, calculated by taking the sum of the returns and dividing by 5, is 14%. However, an investor would actually earn less than 14% per annum by leaving the money invested in the fund for five years. The dollar value of \$100 at the end of the five years would be

$$100 \times 1.15 \times 1.20 \times 1.30 \times 0.80 \times 1.25 = \$179.40$$

By contrast, a 14% return (with annual compounding) would give

$$100 \times 1.14^5 = \$192.54$$

The return that gives \$179.40 at the end of five years is 12.4%. This is because

$$100 \times (1.124)^5 = 179.40$$

What average return should the fund manager report? It is tempting for the manager to make a statement such as: “The average of the returns per year that we have realized in the last five years is 14%.” Although true, this is misleading. It is much less misleading to say: “The average return realized by someone who invested with us for the last five years is 12.4% per year.” In some jurisdictions, regulations require fund managers to report returns the second way.

This phenomenon is an example of a result that is well known by mathematicians. The geometric mean of a set of numbers (not all the same) is always less than the arithmetic mean. In our example, the return multipliers each year are 1.15, 1.20, 1.30, 0.80, and 1.25. The arithmetic mean of these numbers is 1.140, but the geometric mean is only 1.124. An investor who keeps an investment for several years earns a return corresponding to the geometric mean, not the arithmetic mean.

4.4 Regulation

Because they solicit funds from small retail customers, many of whom are unsophisticated, mutual funds and ETFs are heavily regulated. The SEC is the primary regulator of these funds in the United States. The funds must file a registration document with the SEC. Full and accurate financial information must be provided to prospective fund purchasers in a prospectus. There are rules to prevent conflicts of interest, fraud, and excessive fees.

Despite the regulations, there have been a number of scandals involving mutual funds. One of these involves *late trading*. As mentioned earlier in this chapter, if a request to buy or sell mutual fund shares is placed by an investor with a broker by 4 P.M. on any given business day, it is the NAV of the fund at 4 P.M. that determines the price that is paid or received by the investor. In practice, for various reasons, an order to buy or sell is sometimes not passed from a broker to a mutual fund until later than 4 P.M. This allows brokers to collude with investors and submit new orders or change existing orders after 4 P.M. The NAV of the fund at 4 P.M. still applies to the investor—even though he or she may be using information on market movements (particularly movements in overseas markets) after 4 P.M. Late trading is not permitted under SEC regulations, and there were a number of prosecutions in the early 2000s that led to multimillion-dollar fine payments and employees being fired.

Another scandal is known as *market timing*. This is a practice where favored clients are allowed to buy and sell mutual funds shares frequently (e.g., every few days) and in large quantities without penalty. One reason why they might want to do this is that they are indulging in the illegal practice of late trading. Another reason is that they are analyzing the impact of stocks whose prices have not been updated recently on the fund's NAV. Suppose that the price of a stock has not been updated for several hours. (This could be because it does not trade frequently or because it trades on an exchange in a country in a different time zone.) If the U.S. market has gone up (down) in the last few hours, the calculated NAV is likely to understate (overstate) the value of the underlying portfolio and there is a short-term trading opportunity. Taking advantage of this is not necessarily illegal. However, it may be illegal for the mutual fund to offer special trading privileges to favored customers because the costs (such as those associated with providing the liquidity necessary to accommodate frequent redemptions) are borne by all customers.

Other scandals have involved *front running* and *directed brokerage*. Front running occurs when a mutual fund is planning a big trade that is expected to move the market. It informs favored customers or partners before executing the trade, allowing them to trade for their own account first. Directed brokerage involves an improper arrangement between a mutual fund and a brokerage house where the brokerage house recommends the mutual fund to clients in return for receiving orders from the mutual fund for stock and bond trades.

4.5 Hedge Funds

Hedge funds are different from mutual funds in that they are subject to very little regulation. This is because they accept funds only from financially sophisticated individuals and organizations. Examples of the regulations that affect mutual funds are the requirements that:

- Shares be redeemable at any time
- NAV be calculated daily
- Investment policies be disclosed
- The use of leverage be limited

Hedge funds are largely free from these regulations. This gives them a great deal of freedom to develop sophisticated, unconventional, and proprietary investment strategies. Hedge funds are sometimes referred to as *alternative investments*.

The first hedge fund, A. W. Jones & Co., was created by Alfred Winslow Jones in the United States in 1949. It was structured as a general partnership to avoid SEC regulations. Jones combined long positions in stocks considered to be undervalued with short positions in stocks considered to be overvalued. He used leverage to magnify returns. A performance fee equal to 20% of profits was charged to investors. The fund performed well and the term “hedge fund” was coined in a newspaper article written about A. W. Jones & Co. by Carol Loomis in 1966. The article showed that the fund’s performance after allowing for fees was better than the most successful mutual funds. Not surprisingly, the article led to a great deal of interest in hedge funds and their investment approach. Other hedge fund pioneers were George Soros, Walter J. Schloss, and Julian Robertson.⁷

“Hedge fund” implies that risks are being hedged. Jones’s trading strategy did involve hedging. He had little exposure to the overall direction of the market because his long position (in stocks considered to be undervalued) at any given time was about the same size as his short position (in stocks considered to be overvalued). However, for some hedge funds, the word “hedge” is inappropriate because they take aggressive bets on the future direction of the market with no particular hedging policy.

Hedge funds have grown in popularity over the years, and it is estimated that about \$3.8 trillion was invested with them throughout the world in 2021. The success of hedge funds is perhaps a little surprising because, as we will see later, they have performed less well than the S&P 500 between 2009 and 2021. Many hedge funds are registered in tax-favorable jurisdictions such as the Cayman Islands. Funds of funds have been set up to allocate funds to different hedge funds. Hedge funds are difficult to ignore. They account

⁷The famous investor Warren Buffett can also be considered a hedge fund pioneer. In 1956, he started Buffett Partnership LP with seven limited partners and \$100,100. Buffett charged his partners 25% of profits above a hurdle rate of 25%. He searched for unique situations, merger arbitrage, spin-offs, and distressed debt opportunities and earned an average of 29.5% per year. The partnership was disbanded in 1969, and Berkshire Hathaway (a holding company, not a hedge fund) was formed.

for a large part of the daily turnover on the New York and London stock exchanges. They are major players in the convertible bond, credit default swap, distressed debt, and non-investment-grade bond markets. They are also active participants in the ETF market, often taking short positions.

4.5.1 Fees

One characteristic of hedge funds that distinguishes them from mutual funds is that fees are higher and dependent on performance. An annual management fee that is usually between 1% and 3% of assets under management is charged. This is designed to meet operating costs—but there may be an additional fee for things such as audits, account administration, and trader bonuses. Moreover, an incentive fee that is usually between 15% and 30% of realized net profits (i.e., profits after management fees) is charged if the net profits are positive. This fee structure is designed to attract the most talented and sophisticated investment managers. Thus, a typical hedge fund fee schedule might be expressed as “2 plus 20%” indicating that the fund charges 2% per year of assets under management and 20% of net profit. On top of high fees there is usually a lock-up period of at least one year during which invested funds cannot be withdrawn. Some hedge funds with good track records have sometimes charged much more than the average. An example is James Simons’ Renaissance Technologies Corp., which has charged as much as “5 plus 44%.” (James Simons is a former math professor whose wealth was estimated at about \$25 billion in 2021.)

The agreements offered by hedge funds may include clauses that make the incentive fees more palatable. For example:

- There is sometimes a *hurdle rate*. This is the minimum return necessary for the incentive fee to be applicable.
- There is sometimes a *high-water mark clause*. This states that any previous losses must be recouped by new profits before an incentive fee applies. Because different investors place money with the fund at different times, the high-water mark is not necessarily the same for all investors. There may be a *proportional adjustment clause* stating that, if funds are withdrawn by investors, the amount of previous losses that has to be recouped is adjusted proportionally. Suppose a fund worth \$200 million loses \$40 million and \$80 million of funds are withdrawn. The high-water mark clause on its own would require \$40 million of profits on the remaining \$80 million to be achieved before the incentive fee applied. The proportional adjustment clause would reduce this to \$20 million because the fund is only half as big as it was when the loss was incurred.
- There is sometimes a *clawback clause* that allows investors to apply part or all of previous incentive fees to current losses. A portion of the incentive fees paid by the investor each year is then retained in a *recovery account*. This account is used to compensate investors for a percentage of any future losses.

Some hedge fund managers have become very rich from the generous fee schedules. *Forbes* estimates that the top 25 hedge fund managers were worth a combined \$220

billion in 2021. Jim Simons, the math professor mentioned earlier, topped the list. Ray Dalio (Bridgewater Associates) was second.

If an investor has a portfolio of investments in hedge funds, the fees paid can be quite high. As a simple example, suppose that an investment is divided equally between two funds, A and B. Both funds charge 2 plus 20%. In the first year, Fund A earns 20% while Fund B earns −10%. The investor's average return on investment before fees is $0.5 \times 20\% + 0.5 \times (-10\%)$ or 5%. The fees paid to fund A are $2\% + 0.2 \times (20 - 2)\%$ or 5.6%. The fees paid to Fund B are 2%. The average fee paid on the investment in the hedge funds is therefore 3.8%. The investor is left with a 1.2% return. This is half of what the investor would get if 2 plus 20% were applied to the overall 5% return.

When a fund of funds is involved, there is an extra layer of fees and the investor's return after fees is even worse. A typical fee charged by a fund of hedge funds used to be 1% of assets under management plus 10% of the net (after management and incentive fees) profits of the hedge funds they invest in. These fees have gone down as a result of poor hedge fund performance. Suppose a fund of hedge funds divides its money equally between 10 hedge funds. All charge 2 plus 20% and the fund of hedge funds charges 1 plus 10%. It sounds as though the investor pays 3 plus 30%—but it can be much more than this. Suppose that five of the hedge funds lose 40% before fees and the other five make 40% before fees. An incentive fee of 20% of 38% or 7.6% has to be paid to each of the profitable hedge funds. The total incentive fee is therefore 3.8% of the funds invested. In addition there is a 2% annual fee paid to the hedge funds and 1% annual fee paid to the fund of funds. (In this example, the incentive fee to the fund of funds is zero.) The investor's net return is −6.8% of the amount invested. (This is 6.8% less than the return on the underlying assets before fees.)

4.5.2 *Incentives of Hedge Fund Managers*

The fee structure gives hedge fund managers an incentive to make a profit. But it also encourages them to take risks. The hedge fund manager has a call option on the assets of the fund. As is well known, the value of a call option increases as the volatility of the underlying assets increases. This means that the hedge fund manager can increase the value of the option by taking risks that increase the volatility of the fund's assets. The fund manager has a particular incentive to do this when nearing the end of the period over which the incentive fee is calculated and the return to date is low or negative.

Suppose that a hedge fund manager is presented with an opportunity where there is a 0.4 probability of a 60% profit and a 0.6 probability of a 60% loss with the fees earned by the hedge fund manager being 2 plus 20%. The expected return on the investment is

$$0.4 \times 60\% + 0.6 \times (-60\%)$$

or −12%.

Even though this is a terrible expected return, the hedge fund manager might be tempted to accept the investment. If the investment produces a 60% profit, the hedge

Table 4.4 Return from a High-Risk Investment Where Returns of +60% and −60% Have Probabilities of 0.4 and 0.6, Respectively, and the Hedge Fund Charges 2 Plus 20%

Expected return to hedge fund	6.64%
Expected return to investors	−18.64%
Overall expected return	−12.00%

fund's fee is $2 + 0.2 \times 58$ or 13.6%. If the investment produces a 60% loss, the hedge fund's fee is 2%. The expected fee to the hedge fund is therefore

$$0.4 \times 13.6 + 0.6 \times 2 = 6.64$$

or 6.64% of the funds under administration. The expected management fee is 2% and the expected incentive fee is 4.64%.

To the investors in the hedge fund, the expected return is

$$0.4 \times (60 - 0.2 \times 58 - 2) + 0.6 \times (-60 - 2) = -18.64$$

or −18.64%.

The example is summarized in Table 4.4. It shows that the fee structure of a hedge fund gives its managers an incentive to take high risks even when expected returns are negative. The incentives may be reduced by hurdle rates, high–water mark clauses, and clawback clauses. However, these clauses are not always as useful to investors as they appear. One reason is that investors have to continue to invest with the fund to take advantage of them. Another is that, as losses mount for a hedge fund, the hedge fund managers have an incentive to wind up the hedge fund and start a new one.

The incentives we are talking about here are real. Imagine how you would feel as an investor in the hedge fund Amaranth. One of its traders, Brian Hunter, liked to make huge bets on the price of natural gas. Until 2006, his bets were largely right and as a result he was regarded as a star trader. His remuneration including bonuses is reputed to have been close to \$100 million in 2005. During 2006, his bets proved wrong and Amaranth, which had about \$9 billion of assets under administration, lost a massive \$6.5 billion. (This was even more than the loss of hedge fund Long-Term Capital Management in 1998.) Brian Hunter did not have to return the bonuses he had previously earned. Instead, he left Amaranth and tried to start his own hedge fund.

It is interesting to note that, in theory, two individuals can create a money machine as follows. One starts a hedge fund with a certain high risk (and secret) investment strategy. The other starts a hedge fund with an investment strategy that is the opposite of that followed by the first hedge fund. For example, if the first hedge fund decides to buy \$1 million of silver, the second hedge fund shorts this amount of silver. At the time they start the funds, the two individuals enter into an agreement to share the incentive fees. One hedge fund (we do not know which one) is likely to do well and earn good

incentive fees. The other will do badly and earn no incentive fees. Provided that they can find investors for their funds, they have a money machine!

4.5.3 Prime Brokers

Prime brokers are the banks that offer services to hedge funds. Typically a hedge fund, when it is first started, will choose a particular bank as its prime broker. This bank handles the hedge fund's trades (which may be with the prime broker or with other financial institutions), carries out calculations each day to determine the collateral the hedge fund has to provide, borrows securities for the hedge fund when it wants to take short positions, provides cash management and portfolio reporting services, and makes loans to the hedge fund. In some cases, the prime broker provides risk management and consulting services and introduces the hedge fund to potential investors. The prime broker has a good understanding of the hedge fund's portfolio and will typically carry out stress tests on the portfolio to decide how much leverage it is prepared to offer the fund.

Although hedge funds are not heavily regulated, they do have to answer to their prime brokers. The prime broker is the main source of borrowed funds for a hedge fund. The prime broker monitors the risks being taken by the hedge fund and determines how much the hedge fund is allowed to borrow. Typically a hedge fund has to post securities with the prime broker as collateral for its loans. When it loses money, more collateral has to be posted. If it cannot post more collateral, it has no choice but to close out its trades. One thing the hedge fund has to think about is the possibility that it will enter into a trade that is correct in the long term, but loses money in the short term. Consider a hedge fund that thinks credit spreads are too high. It might be tempted to take a highly leveraged position where BBB-rated bonds are bought and Treasury bonds are shorted. However, there is the danger that credit spreads will increase before they decrease. In this case, the hedge fund might run out of collateral and be forced to close out its position at a huge loss.

As a hedge fund gets larger, it is likely to use more than one prime broker. This means that no one bank sees all its trades and has a complete understanding of its portfolio. The opportunity of transacting business with more than one prime broker gives a hedge fund more negotiating clout to reduce the fees it pays. Goldman Sachs, Morgan Stanley, and many other large banks offer prime broker services to hedge funds and find them to be an important contributor to their profits.⁸

⁸Although a bank is taking some risks when it lends to a hedge fund, it is also true that a hedge fund is taking some risks when it chooses a prime broker. Many hedge funds that chose Lehman Brothers as their prime broker found that they could not access assets, which they had placed with Lehman Brothers as collateral, when the company went bankrupt in 2008.

4.6 Hedge Fund Strategies

In this section we will discuss the strategies followed by hedge funds. Our classification is similar to the one used by the Barclay Hedge Fund Indices, which provides indices tracking hedge fund performance. Not all hedge funds can be classified in the way indicated. Some follow more than one of the strategies mentioned and some follow strategies that are not listed. (For example, there are funds specializing in weather derivatives.)

4.6.1 Long/Short Equity

As described earlier, long/short equity strategies were used by hedge fund pioneer Alfred Winslow Jones. They continue to be among the most popular of hedge fund strategies. The hedge fund manager identifies a set of stocks that are considered to be undervalued by the market and a set that are considered to be overvalued. The manager takes a long position in the first set and a short position in the second set. Typically, the hedge fund has to pay the prime broker a fee (perhaps 1% per year) to rent the shares that are borrowed for the short position. (See Chapter 5 for a discussion of short selling.)

Long/short equity strategies are all about stock picking. If the overvalued and undervalued stocks have been picked well, the strategies should give good returns in both bull and bear markets. Hedge fund managers often concentrate on smaller stocks that are not well covered by analysts and research the stocks extensively using fundamental analysis, as pioneered by Benjamin Graham. The hedge fund manager may choose to maintain a net long bias where the shorts are of smaller magnitude than the longs or a net short bias where the reverse is true. Alfred Winslow Jones maintained a net long bias in his successful use of long/short equity strategies.

An *equity-market-neutral* fund is one where longs and shorts are matched in some way. A *dollar-neutral* fund is an equity-market-neutral fund where the dollar amount of the long position equals the dollar amount of the short position. A *beta-neutral fund* is a more sophisticated equity-market-neutral fund where the weighted average beta of the shares in the long portfolio equals the weighted average beta of the shares in the short portfolio so that the overall beta of the portfolio is zero. If the capital asset pricing model is true, the beta-neutral fund should be totally insensitive to market movements. Long and short positions in index futures are sometimes used to maintain a beta-neutral position.

Sometimes equity-market-neutral funds go one step further. They maintain *sector neutrality* where long and short positions are balanced by industry sectors or *factor neutrality* where the exposure to factors such as the price of oil, the level of interest rates, or the rate of inflation is neutralized.

4.6.2 Dedicated Short

Managers of dedicated short funds look exclusively for overvalued companies and sell them short. They are attempting to take advantage of the fact that brokers and analysts are reluctant to issue sell recommendations—even though one might reasonably

expect the number of companies overvalued by the stock market to be approximately the same as the number of companies undervalued at any given time. Typically, the companies chosen are those with weak financials, those that change their auditors regularly, those that delay filing reports with the SEC, companies in industries with overcapacity, companies suing or attempting to silence their short sellers, and so on.

4.6.3 Distressed Securities

Bonds with credit ratings of BB or lower are known as “non-investment-grade” or “junk” bonds. Those with a credit rating of CCC are referred to as “distressed” and those with a credit rating of D are in default. Typically, distressed bonds sell at a big discount to their par value and provide a yield that is over 1,000 basis points (10%) more than the yield on Treasury bonds. Of course, an investor only earns this yield if the required interest and principal payments are actually made.

The managers of funds specializing in distressed securities carefully calculate a fair value for distressed securities by considering possible future scenarios and their probabilities. Distressed debt cannot usually be shorted and so they search for debt that is undervalued by the market. Bankruptcy proceedings usually lead to a reorganization or liquidation of a company. The fund managers understand the legal system, know priorities in the event of liquidation, estimate recovery rates, consider actions likely to be taken by management, and so on.

Some funds are passive investors. They buy distressed debt when the price is below its fair value and wait. Other hedge funds adopt an active approach. They might purchase a sufficiently large position in outstanding debt claims so that they have the right to influence a reorganization proposal. In Chapter 11 reorganizations in the United States, each class of claims must approve a reorganization proposal with a two-thirds majority. This means that one-third of an outstanding issue can be sufficient to stop a reorganization proposed by management or other stakeholders. In a reorganization of a company, the equity is often worthless and the outstanding debt is converted into new equity. Sometimes, the goal of an active manager is to buy more than one-third of the debt, obtain control of a target company, and then find a way to extract wealth from it.

4.6.4 Merger Arbitrage

Merger arbitrage involves trading after a merger or acquisition is announced in the hope that the announced deal will take place. There are two main types of deals: cash deals and share-for-share exchanges.

Consider first cash deals. Suppose that Company A announces that it is prepared to acquire all the shares of Company B for \$30 per share. Suppose the shares of Company B were trading at \$20 prior to the announcement. Immediately after the announcement its share price might jump to \$28. It does not jump immediately to \$30 because (a) there is some chance that the deal will not go through and (b) it may take some time for the full impact of the deal to be reflected in market prices. Merger-arbitrage hedge funds

buy the shares in company B for \$28 and wait. If the acquisition goes through at \$30, the fund makes a profit of \$2 per share. If it goes through at a higher price, the profit is higher. However, if for any reason the deal does not go through, the hedge fund will take a loss.

Consider next a share-for-share exchange. Suppose that Company A announces that it is willing to exchange one of its shares for four of Company B's shares. Assume that Company B's shares were trading at 15% of the price of Company A's shares prior to the announcement. After the announcement, Company B's share price might rise to 22% of Company A's share price. A merger-arbitrage hedge fund would buy a certain amount of Company B's stock and at the same time short a quarter as much of Company A's stock. This strategy generates a profit if the deal goes ahead at the announced share-for-share exchange ratio or one that is more favorable to Company B.

Merger-arbitrage hedge funds can generate steady, but not stellar, returns. It is important to distinguish merger arbitrage from the activities of Ivan Boesky and others who used inside information to trade before mergers became public knowledge.⁹ Trading on inside information is illegal. Ivan Boesky was sentenced to three years in prison and fined \$100 million.

4.6.5 Convertible Arbitrage

Convertible bonds are bonds that can be converted into the equity of the bond issuer at certain specified future times with the number of shares received in exchange for a bond possibly depending on the time of the conversion. The issuer usually has the right to call the bond (i.e., buy it back for a prespecified price) in certain circumstances. Usually, the issuer announces its intention to call the bond as a way of forcing the holder to convert the bond into equity immediately. (If the bond is not called, the holder is likely to postpone the decision to convert it into equity for as long as possible.)

A convertible arbitrage hedge fund has typically developed a sophisticated model for valuing convertible bonds. The convertible bond price depends in a complex way on the price of the underlying equity, its volatility, the level of interest rates, and the chance of the issuer defaulting. Many convertible bonds trade at prices below their fair value. Hedge fund managers buy the bond and then hedge their risks by shorting the stock. (This is an application of delta hedging, which will be discussed in Chapter 15.) Interest rate risk and credit risk can be hedged by shorting nonconvertible bonds that are issued by the company that issued the convertible bond. Alternatively, the managers can take positions in interest rate futures contracts, asset swaps, and credit default swaps to accomplish this hedging.

⁹The Michael Douglas character of Gordon Gekko in the award-winning movie *Wall Street* was based on Ivan Boesky.

4.6.6 Fixed-Income Arbitrage

The basic tool of fixed-income trading is the zero-coupon yield curve, the construction of which is discussed in Appendix B. One strategy followed by hedge fund managers that engage in fixed-income arbitrage is a *relative value* strategy, where they buy bonds that the zero-coupon yield curve indicates are undervalued by the market and sell bonds that it indicates are overvalued. *Market-neutral* strategies are similar to relative value strategies except that the hedge fund manager tries to ensure that the fund has no exposure to interest rate movements.

Some fixed-income hedge fund managers follow directional strategies where they take a position based on a belief that a certain spread between interest rates, or interest rates themselves, will move in a certain direction. Usually they have a lot of leverage and have to post collateral. They are therefore taking the risk that they are right in the long term, but that the market moves against them in the short term so that they cannot post collateral and are forced to close out their positions at a loss. This is what happened to Long-Term Capital Management (see Business Snapshot 16.1).

4.6.7 Emerging Markets

Emerging market hedge funds specialize in investments associated with developing countries. Some of these funds focus on equity investments. They screen emerging market companies looking for shares that are overvalued or undervalued. They gather information by traveling, attending conferences, meeting with analysts, talking to management, and employing consultants. Usually they invest in securities trading on the local exchange, but sometimes they use American Depositary Receipts (ADRs). ADRs are certificates issued in the United States and traded on a U.S. exchange. They are backed by shares of a foreign company. ADRs may have better liquidity and lower transaction costs than the underlying foreign shares. Sometimes there are price discrepancies between ADRs and the underlying shares giving rise to arbitrage opportunities.

Another type of investment is debt issued by an emerging market country. Eurobonds are bonds issued by the country and denominated in a hard currency such as the U.S. dollar or the euro. Local currency bonds are bonds denominated in the local currency. Hedge funds invest in both types of bonds. They can be risky: countries such as Russia, Argentina, Brazil, and Venezuela have defaulted several times on their debt.

4.6.8 Global Macro

Global macro is the hedge fund strategy used by star managers such as George Soros and Julian Robertson. Global macro hedge fund managers carry out trades that reflect global macroeconomic trends. They look for situations where markets have, for whatever reason, moved away from equilibrium and place large bets that they will move back into equilibrium. Often the bets are on exchange rates and interest rates. A global macro

strategy was used in 1992 when George Soros's Quantum Fund gained \$1 billion by betting that the British pound would decrease in value. More recently, hedge funds have (with mixed results) placed bets that the huge U.S. balance of payments deficit would cause the value of the U.S. dollar to decline. The main problem for global macro funds is that they do not know when equilibrium will be restored. World markets can for various reasons be in disequilibrium for long periods of time.

4.6.9 *Managed Futures*

Hedge fund managers that use managed futures strategies attempt to predict future movements in commodity prices. Some rely on the manager's judgment; others use computer programs to generate trades. Some managers base their trading on technical analysis, which analyzes past price patterns to predict the future. Others use fundamental analysis, which involves calculating a fair value for the commodity from economic, political, and other relevant factors.

When technical analysis is used, trading rules are usually first tested on historical data. This is known as back-testing. If (as is often the case) a trading rule has come from an analysis of past data, trading rules should be tested out of sample (that is, on data that are different from the data used to generate the rules). Analysts should be aware of the perils of data mining. Suppose thousands of different trading rules are generated and then tested on historical data. Just by chance a few of the trading rules will perform very well—but this does not mean that they will perform well in the future.

4.7 Hedge Fund Performance

It is not as easy to assess hedge fund performance as it is to assess mutual fund performance. There is no data set that records the returns of all hedge funds. For the Lipper hedge funds database (TASS), which is available to researchers, participation by hedge funds is voluntary. Small hedge funds and those with poor track records often do not report their returns and are therefore not included in the data set. When returns are reported by a hedge fund, the database is usually backfilled with the fund's previous returns. This creates a bias in the returns that are in the data set because, as just mentioned, the hedge funds that decide to start providing data are likely to be the ones doing well.

Arguably, hedge funds can improve the risk-return trade-offs available to pension plans. This is because pension plans cannot (or choose not to) take short positions, obtain leverage, invest in derivatives, and engage in many of the complex trades that are favored by hedge funds. Investing in a hedge fund is a simple way in which a pension fund can (for a fee) expand the scope of its investing. This may improve its efficient frontier. (See Section 1.2 for a discussion of efficient frontiers.)

It is not uncommon for hedge funds to report good returns for a few years and then "blow up." Long-Term Capital Management reported returns (before fees) of 28%, 59%, 57%, and 17% in 1994, 1995, 1996, and 1997, respectively. In 1998, it lost virtually all

Table 4.5 Performance of Hedge Funds

Year	Barclays Hedge Fund Index Net Return (%)	S&P 500 Return Including Dividends (%)
2008	−21.63	−37.00
2009	23.74	26.46
2010	10.88	15.06
2011	−5.48	2.11
2012	8.25	16.00
2013	11.12	32.39
2014	2.88	13.39
2015	0.04	1.38
2016	6.10	11.96
2017	10.36	21.83
2018	−5.23	−4.38
2019	10.64	31.49
2020	11.14	18.40
2021	10.20	28.71

its capital. Some people have argued that hedge fund returns are like the returns from selling out-of-the-money options. Most of the time, the options cost nothing, but every so often they are very expensive.

This may be unfair. Advocates of hedge funds would argue that hedge fund managers search for profitable opportunities and use complex analyses that other investors do not have the resources or expertise to find. They would point out that some hedge fund managers have been very successful.

Prior to 2008, hedge funds performed quite well. In 2008, hedge funds on average lost money but provided a better performance than the S&P 500. During the years 2009 to 2021, the S&P 500 provided a much better return than the average hedge fund.¹⁰ The Barclays hedge fund index is the arithmetic average return (net after fees) of all hedge funds (excluding funds of funds) in the Barclays database. (Potentially, it has some of the biases mentioned earlier.) Table 4.5 compares returns given by the index with total returns from the S&P 500.

Summary

Mutual funds and ETFs offer a way small investors can capture the benefits of diversification. Overall, the evidence is that actively managed funds do not outperform the market and this has led many investors to choose funds that are designed to track a market index such as the S&P 500.

¹⁰In fairness, it should be pointed out that hedge funds often have a beta less than one (for example, long/short equity funds are often designed to have a beta close to zero), so a return less than the S&P 500 during periods when the market does very well does not necessarily indicate a negative alpha.

Most mutual funds are open-end funds, so that the number of shares in the fund increases (decreases) as investors contribute (withdraw) funds. An open-end mutual fund calculates the net asset value of shares in the fund at 4 P.M. each business day and this is the price used for all buy and sell orders placed in the previous 24 hours. A closed-end fund has a fixed number of shares that trade in the same way as the shares of any other corporation.

Exchange-traded funds (ETFs) are proving to be popular alternatives to open- and closed-end mutual funds. The assets held by the fund are known at any given time. Large institutional investors can exchange shares in the fund at any time for the assets, and vice versa. This ensures that the shares in the ETF (unlike shares in a closed-end fund) trade at a price very close to the fund's net asset value. Shares in an ETF can be traded at any time (not just at 4 P.M.) and shares in an ETF (unlike shares in an open-end mutual fund) can be shorted.

Hedge funds cater to the needs of large investors. Compared to mutual funds, they are subject to very few regulations and restrictions. Hedge funds charge investors much higher fees than mutual funds. The fee for a typical fund is "2 plus 20%." This means that the fund charges a management fee of 2% per year and receives 20% of the profit after management fees have been paid generated by the fund if this is positive. Hedge fund managers have a call option on the assets of the fund and, as a result, may have an incentive to take high risks.

Among the strategies followed by hedge funds are long/short equity, dedicated short, distressed securities, merger arbitrage, convertible arbitrage, fixed-income arbitrage, emerging markets, global macro, and managed futures. The jury is still out on whether hedge funds provide better risk-return trade-offs than index funds after fees. There is an unfortunate tendency for hedge funds to provide excellent returns for a number of years and then report a disastrous loss.

Further Reading

- Jensen, M. C. "Risk, the Pricing of Capital Assets, and the Evaluation of Investment Portfolios." *Journal of Business* 42, no. 2 (April 1969): 167–247.
- Khorana, A., H. Servaes, and P. Tufano. "Mutual Fund Fees Around the World." *Review of Financial Studies* 22 (March 2009): 1279–1310.
- Lhabitant, F.-S. *Handbook of Hedge Funds*. Chichester, UK: John Wiley & Sons, 2006.
- Ross, S. "Neoclassical Finance, Alternative Finance and the Closed End Fund Puzzle." *European Financial Management* 8 (2002): 129–137.

Practice Questions and Problems (Answers at End of Book)

- 4.1 What is the difference between an open-end and closed-end mutual fund?
- 4.2 How is the NAV of an open-end mutual fund calculated? When is it calculated?

- 4.3 What are the two NAVs of a closed-end mutual fund? Are they usually the same?
- 4.4 What is an index fund? How is it created?
- 4.5 What is a mutual fund's (a) front-end load and (b) back-end load?
- 4.6 Explain how an exchange-traded fund that tracks an index works. What are the advantages of an exchange-traded fund over (a) an open-end mutual fund and (b) a closed-end mutual fund?
- 4.7 What is the difference between the geometric mean and the arithmetic mean of a set of numbers? Why is the difference relevant to the reporting of mutual fund returns?
- 4.8 Explain the meaning of (a) late trading, (b) market timing, (c) front running, and (d) directed brokerage.
- 4.9 Give three examples of the rules that apply to mutual funds, but not to hedge funds.
- 4.10 "If 70% of convertible bond trading is by hedge funds, I would expect the profitability of that strategy to decline." Discuss this viewpoint.
- 4.11 Explain the meanings of the terms hurdle rate, high-water mark clause, and claw-back clause when used in connection with the incentive fees of hedge funds.
- 4.12 A hedge fund charges 2 plus 20%. Investors want a return after fees of 20%. How much does the hedge fund have to earn, before fees, to provide investors with this return? Assume that the incentive fee is paid on the net return after management fees have been subtracted.
- 4.13 "It is important for a hedge fund to be right in the long term. Short-term gains and losses do not matter." Discuss this statement.
- 4.14 "The risks that hedge funds take are regulated by their prime brokers." Discuss this statement.

Further Questions

- 4.15 Good years are followed by equally bad years for a mutual fund. It earns +8%, -8%, +12%, -12% in successive years. What is the overall return for the four years?
- 4.16 A fund of funds divides its money between five hedge funds that earn -5%, 1%, 10%, 15%, and 20% before fees in a particular year. The fund of funds charges 1 plus 10% and the hedge funds charge 2 plus 20%. The hedge funds' incentive fees are calculated on the return after management fees. The fund of funds incentive fee is calculated on the net (after management and incentive fees) average return of the hedge funds in which it invests and after its own management fee has been subtracted. What is the overall return on the investments? How is it divided between the fund of funds, the hedge funds, and investors in the fund of funds?
- 4.17 A hedge fund charges 2 plus 20%. A pension fund invests in the hedge fund. Plot the return to the pension fund as a function of the return to the hedge fund.

Part Two

FINANCIAL MARKETS

Chapter 5

Financial Instruments

Financial institutions do a huge volume of trading in a wide range of different financial instruments. There are a number of reasons for this. Some trades are designed to satisfy the needs of their clients, some are to manage their own risks, some are to exploit arbitrage opportunities, and some are to reflect their own views on the direction in which market prices will move. (The Volcker rule in the Dodd–Frank Act, which will be discussed in Chapter 26, restricts the extent to which banks can engage in speculative trading.)

In this chapter, we discuss different financial instruments and how they trade. There are two main markets for trading financial instruments. These are the *exchange-traded market* and the *over-the-counter* (or OTC) *market*. The OTC market is discussed in the next chapter. At this stage we note that exchanges define the contracts that can be traded and organize trading to ensure that contracts are honored. The OTC market links traders who work for financial institutions, large corporations, and fund managers. Trades are typically much larger than those on an exchange and the terms of a trade do not have to be those specified by exchanges.

5.1 Long and Short Positions in Assets

The simplest type of trade is the purchase of an asset for cash or the sale of an asset that is owned for cash. Examples of such trades are:

- The purchase of 100 IBM shares
- The sale of 1 million British pounds for U.S. dollars

- The purchase of 1,000 ounces of gold
- The sale of \$1 million worth of bonds issued by General Motors

The first of these trades would typically be done on an exchange; the other three would be done in the OTC market. The trades are sometimes referred to as *spot trades* because they lead to almost immediate (on the spot) delivery of the asset.

When using a broker to purchase an asset on an exchange, a trader can often borrow part of the cost of the asset from the broker, pledging the asset as collateral. This is known as *buying on margin*. The broker's risk is that the price of the asset will decline sharply. Accordingly, the broker monitors the margin account balance. This is the part of the purchase price paid by the investor adjusted for gains or losses on the asset. If the balance in the margin account falls below a certain level, known as the maintenance margin, the trader is required to bring the balance back up to that level. Suppose a trader buys 1,000 shares at \$120 per share on margin. Suppose further the amount borrowed is 50% of the price of the asset and the maintenance margin is 25% of the price of the asset. The trader will have to provide half the cost or \$60,000. If the share price falls to \$78, there is a \$42,000 loss and the balance in the margin account becomes \$18,000. This is $18,000/78,000$ or 23.1% of the value of the asset, and the trader will have to provide \$1,500 of cash to bring the balance in the margin account up to 25% of \$78,000. If the trader fails to do this, the broker sells the position. This example corresponds to the limits of what is allowed in the United States. Regulations allow an investor to finance up to 50% of the price by buying on margin. The minimum maintenance margin is 25% of the price of the asset.

5.1.1 Short Sales

In some markets, it is possible to sell an asset that you do not own with the intention of buying it back later. This is referred to as shorting the asset. We will illustrate how it works by considering the shorting of shares of a stock.

Suppose an investor instructs a broker to short 500 shares of a stock that is traded on an exchange. The broker will carry out the instructions by borrowing the shares from another client and selling them on an exchange in the usual way. (A small fee may be charged for the borrowed shares.) The investor can maintain the short position for as long as desired, provided there are always shares available for the broker to borrow. At some stage, however, the investor will close out the position by purchasing 500 shares. These are then replaced in the account of the client from whom the shares were borrowed. The investor takes a profit if the stock price has declined and a loss if it has risen. If, at any time while the contract is open, the broker is unable to borrow shares, the investor is *short-squeezed* and is forced to close out the position immediately, even if not ready to do so.

An investor with a short position must pay to the broker any income, such as dividends or interest, that would normally be received on the asset that has been shorted. The

Table 5.1 Cash Flows from Short Sale and Purchase of Shares

Purchase of Shares	
April: Purchase 500 shares for \$120	−\$60,000
May: Receive dividend	+\$500
July: Sell 500 shares for \$100 per share	+\$50,000
Net Profit = −\$9,500	
Short Sale of Shares	
April: Borrow 500 shares and sell them for \$120	+\$60,000
May: Pay dividend	−\$500
July: Buy 500 shares for \$100 per share	−\$50,000
Replace borrowed shares to close short position	
Net Profit = +\$9,500	

broker will transfer this to the client account from which the asset was borrowed. Suppose a trader shorts 500 shares in April when the price per share is \$120 and closes out the position by buying them back in July when the price per share is \$100. Suppose further that a dividend of \$1 per share is paid in May. The investor receives $500 \times \$120 = \$60,000$ in April when the short position is initiated. The dividend leads to a payment by the investor of $500 \times \$1 = \500 in May. The investor also pays $500 \times \$100 = \$50,000$ for shares when the position is closed out in July. The net gain is, therefore,

$$\$60,000 - \$500 - \$50,000 = \$9,500$$

Table 5.1 illustrates this example and shows that (assuming no fee is charged for borrowing the shares) the cash flows from the short sale are the mirror image of the cash flows from purchasing the shares in April and selling them in July.

An investor entering into a short position has potential future liabilities and is therefore required to provide margin (i.e., collateral). In the United States, the margin is typically 150% of the value of the shares shorted. The proceeds of the sale typically form part of the margin account, so an extra 50% of the value of the shares must be provided.

Suppose that 100 shares are shorted at \$60. The margin required is 150% of the value of the shares or \$9,000. The proceeds of the sale provide \$6,000, so the trader must provide an extra \$3,000. Suppose that a maintenance margin of 125% is set, so that when the balance in the margin account falls below 125% of the value of the shares there is a margin call requiring that the balance be brought up to this level. If the share price rises to \$80, 125% of the value of the shares shorted is \$10,000, and an extra \$1,000 of margin would be required. The short trader must provide this to avoid having the short position closed out.

From time to time, regulations are changed on short selling. Prior to 2007, an *uptick rule* existed in the United States. This meant that stocks could only be shorted when the most recent price movement had been up. The rule was designed to stop short sellers driving down prices. In 2010, after the financial crisis of 2008, an *alternative uptick rule* was introduced. This states that the uptick rule applies in circumstances where there has

been a decline of more than 10% in the price of a stock in one day. The rule applies for the rest of that day and the next one. Occasionally, there are temporary bans on short selling. For example, on September 19, 2008, in an attempt to halt the slide in bank stock prices, the SEC imposed a temporary ban on the short selling of 799 financial companies. This was similar to a ban imposed by the UK Financial Services Authority (FSA) the day before.

5.2 Derivatives Markets

A derivative is an instrument whose value depends on (or derives from) other more basic market variables. A stock option, for example, is a derivative whose value is dependent on the price of a stock.

Derivatives trade in both the exchange-traded and OTC markets. Both markets are huge. Although the statistics that are collected for the two markets are not exactly comparable, it is clear that the over-the-counter derivatives market is much larger than the exchange-traded derivatives market. The Bank for International Settlements (www.bis.org) collects statistics on the markets. Figure 5.1 compares (a) the estimated total principal amounts underlying transactions that were outstanding in the over-the-counter markets between June 1998 and December 2021 and (b) the estimated total value of the assets underlying exchange-traded contracts during the same period. Using these measures, in December 2021, the size of the over-the-counter market was \$598.4 trillion and the size of the exchange-traded market was \$80.09 trillion. Figure 5.1 shows that the OTC market grew rapidly until 2007, but has seen little net growth since then. The

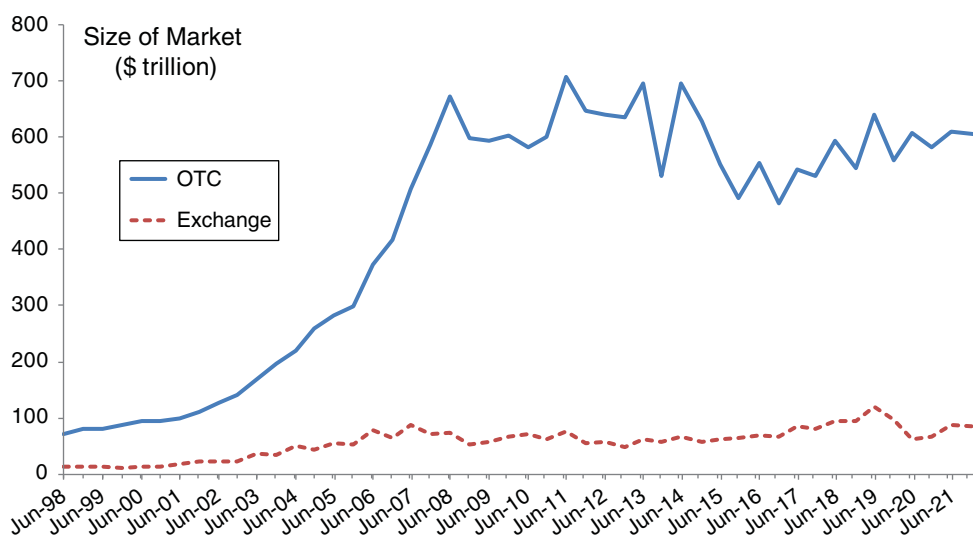


Figure 5.1 Size of Over-the-Counter and Exchange-Traded Derivatives Markets between 1998 and 2021

decline of the market in 2014 and 2015 was largely a result of *compression*. This is a procedure where two or more counterparties restructure transactions with each other with the result that the underlying principal is reduced.

In interpreting these numbers, we should bear in mind that the principal (or value of assets) underlying a derivatives transaction is not the same as its value. An example of an over-the-counter contract could be an agreement entered into some time ago to buy 100 million U.S. dollars with British pounds at a predetermined exchange rate in one year. The total principal amount underlying this transaction is \$100 million. However, the value of the contract might be only \$1 million. The Bank for International Settlements estimated the gross market value of all OTC contracts outstanding in December 2021 to be about \$12.4 trillion.

5.3 Plain Vanilla Derivatives

This section discusses the standard, or commonly traded, contracts in derivatives markets: forwards, futures, swaps, and options. They are sometimes referred to as *plain vanilla* products.

5.3.1 Forward Contracts

A forward contract is an agreement to buy an asset in the future for a certain price. Forward contracts trade in the over-the-counter market. One of the parties to a forward contract assumes a *long position* and agrees to buy the underlying asset on a certain specified future date for a certain specified price. The other party assumes a *short position* and agrees to sell the asset on the same date for the same price.

Forward contracts on foreign exchange are very popular. Table 5.2 provides quotes on the exchange rate between the British pound (GBP) and the U.S. dollar (USD) that might be provided by a large international bank in April 2022. The quotes are for the number of USD per GBP. The first row indicates that the bank is prepared to buy GBP (also known as sterling) in the spot market (i.e., for virtually immediate delivery) at the rate of \$1.2732 per GBP and sell sterling in the spot market at \$1.2736 per GBP; the second row indicates that the bank is prepared to buy sterling in one month at \$1.2746 per GBP and sell sterling in one month at \$1.2751 per GBP; and so on.

Table 5.2 Spot and Forward Quotes for the USD/GBP Exchange
Rate (GBP = British pound; USD = U.S. dollar; quote is number of
USD per GBP)

	Bid	Ask
Spot	1.2732	1.2736
1-month forward	1.2746	1.2751
3-month forward	1.2772	1.2777
1-year forward	1.2883	1.2889

Forward contracts can be used to hedge foreign currency risk. Suppose that at the time of the quotes, the treasurer of a U.S. corporation knows that the corporation will pay £1 million in one year and wants to hedge against exchange rate moves. The treasurer can agree to buy £1 million one-year forward at an exchange rate of 1.2889 by trading with the bank providing the quotes in Table 5.2. The corporation then has a long forward contract on GBP. It has agreed that in one year it will buy £1 million from the bank for \$1.2889 million. The bank has a short forward contract on GBP. It has agreed that in one year it will sell £1 million for \$1.2889 million. Both the corporation and the bank have made a binding commitment.

What are the possible outcomes in the trade we have just described? The forward contract obligates the corporation to purchase 1 million pounds at an exchange rate of 1.2889 in one year. If the spot exchange rate applicable to the corporation when buying pounds rose to, say, 1.5000, at the end of one year the forward contract leads to 1 million pounds being purchased by the corporation at an exchange rate of 1.2889 rather than at the spot exchange rate of 1.5000. This is worth $\$211,100 = (1.5000 - 1.2889) \times \$1,000,000$ to the corporation. Of course, it may also happen that the contract will have a negative final value to the corporation. If the exchange rate falls to 1.1000 by the end of the year, the forward contract leads to 1 million pounds being purchased by the corporation at an exchange rate of 1.2889 rather than at the 1.1000 available in the market. This costs the corporation $\$188,900 = (1.2889 - 1.1000) \times \$1,000,000$. This example shows that a long forward contract can lead to a payoff that is a gain or loss. The payoff is the spot price of the assets underlying the forward contract minus the agreed delivery price for the assets and is shown in Figure 5.2(a).

The bank in our example has entered into a short forward contract. Its position is the mirror image of that of the corporation. The bank has agreed to sell 1 million pounds for an exchange rate of 1.2889 in one year. If the spot exchange rate applicable to the bank rose to 1.5000, at the end of the year the forward contract leads to 1 million pounds being sold by the bank at an exchange rate of 1.2889 rather than at the spot exchange rate of 1.5000. This costs the bank $\$211,100$. If the exchange rate falls to 1.1000 by the end of the year, the forward contract leads to 1 million pounds being sold by the bank at an exchange rate of 1.2889 rather than 1.1000. This is worth $\$188,900$ to the bank. The payoff is the agreed delivery price for the assets underlying the forward contract minus the spot price and is shown in Figure 5.2(b). The valuation of forward contracts and the determination of forward prices are discussed in Appendix C.

5.3.2 *Futures Contracts*

Futures contracts, like forward contracts, are agreements to buy an asset at a future time. Unlike forward contracts, futures are traded on an exchange. This means that the contracts that trade are standardized. The exchange defines the amount of the asset underlying one contract, when delivery can be made, exactly what can be delivered, and so on. Contracts are referred to by their delivery month. For example, the September 2024 gold futures is a contract where delivery is made in September 2024. Whereas

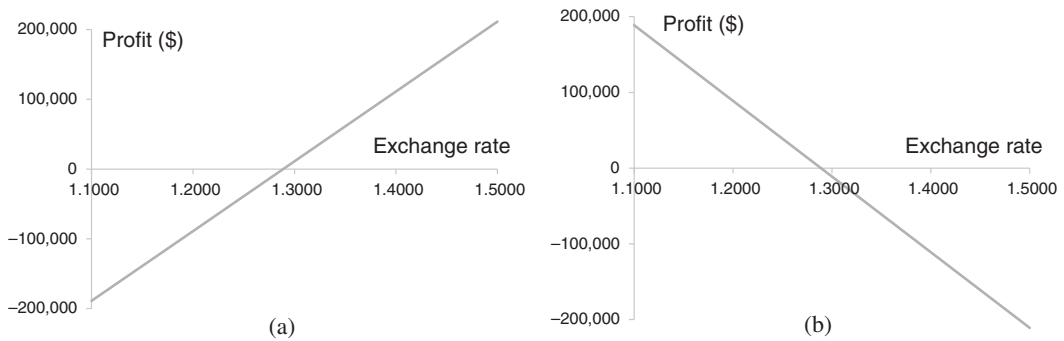


Figure 5.2 Payoffs from Forward Contracts (a) Long position to buy 1 million British pounds in one year. (b) Short position to sell 1 million British pounds in one year. Forward price = 1.2889.

only one delivery day is usually specified for a forward contract, there is usually a period of time during the delivery month when delivery can take place in a futures contract. Alternative delivery times, delivery locations, and so on, are defined by the exchange. It is nearly always the party with the short position that has the right to initiate the delivery process and choose between the alternatives. As we will explain shortly, futures traders have potential future liabilities and are required to post margin.

Most futures contracts trade actively with the futures price at any given time being determined by supply and demand. If there are more buyers than sellers at a time when the September 2024 future price of gold is \$1,280 per ounce, the price goes up. Similarly, if there are more sellers than buyers, the price goes down.

One of the attractive features of exchange-traded contracts such as futures is that it is easy to close out a position. If you buy (i.e., take a long position in) a September 2024 gold futures contract on March 5, 2024, you can exit on June 5, 2024, by selling (i.e., taking a short position in) the same contract. Closing out a position in a forward contract is not as easy as closing out a position in a futures contract. As a result, forward contracts usually lead to final delivery of the underlying assets, whereas futures contracts are usually closed out before the delivery month is reached. Business Snapshot 5.1 is an amusing story showing that the assets underlying futures contracts do get delivered if mistakes are made in the close out.

The futures price of an asset is usually very similar to its forward price. Appendix C at the end of the book gives the relationship between the futures or forward price of an asset and its spot price. One difference between a futures and a forward contract is that a futures is settled daily whereas a forward is settled at the end of its life. For example, if a futures price increases during a day, money flows from traders with short positions to traders with long positions at the end of the day. Similarly, if a futures price decreases during a day, money flows in the opposite direction. Because a futures contract is settled daily whereas a forward contract is settled at the end of its life, the timing of the realization of gains and losses is different for the two contracts. This sometimes causes confusion, as indicated in Business Snapshot 5.2. Table 5.3 summarizes the difference between forward and futures contracts.

BUSINESS SNAPSHOT 5.1

The Unanticipated Delivery of a Futures Contract

This story (which may well be apocryphal) was told to the author of this book many years ago by a senior executive of a financial institution. It concerns a new employee of the financial institution who had not previously worked in the financial sector. One of the clients of the financial institution regularly entered into a long futures contract on live cattle for hedging purposes and issued instructions to close out the position on the last day of trading. (Live cattle futures contracts trade on the Chicago Mercantile Exchange and each contract is on 40,000 pounds of cattle.) The new employee was given responsibility for handling the account.

When the time came to close out a contract, the employee noted that the client was long one contract and instructed a trader at the exchange go long (not short) one contract. The result of this mistake was that the financial institution ended up with a long position in two live cattle futures contracts. By the time the mistake was spotted, trading in the contract had ceased.

The financial institution (not the client) was responsible for the mistake. As a result it started to look into the details of the delivery arrangements for live cattle futures contracts—something it had never done before. Under the terms of the contract, cattle could be delivered by the party with the short position to a number of different locations in the United States during the delivery month. Because it was long, the financial institution could do nothing but wait for a party with a short position to issue a *notice of intention to deliver* to the exchange and for the exchange to assign that notice to the financial institution.

It eventually received a notice from the exchange and found that it would receive live cattle at a location 2,000 miles away the following Tuesday. The new employee was dispatched to the location to handle things. It turned out that the location had a cattle auction every Tuesday. The party with the short position that was making delivery bought cattle at the auction and then immediately delivered them. Unfortunately, the cattle could not be resold until the next cattle auction the following Tuesday. The employee was therefore faced with the problem of making arrangements for the cattle to be housed and fed for a week. This was a great start to a first job in the financial sector!

BUSINESS SNAPSHOT 5.2

A Software Error?

A foreign exchange trader working for a bank enters into a long forward contract to buy 1 million pounds sterling at an exchange rate of 1.3000 in three months. At the same time, another trader on the next desk takes a long position in 16 three-month futures contracts on sterling. The futures price is 1.3000 and each contract is on 62,500 pounds. The sizes of the positions taken by the forward and futures traders are therefore the same. Within minutes of the positions being taken, the forward and the futures prices both increase to 1.3040. The bank's systems show that the futures trader has made a profit of \$4,000 while the forward trader has made a profit of only \$3,900. The forward trader immediately calls the bank's systems department to complain. Does the forward trader have a valid complaint?

The answer is no! The daily settlement of futures contracts ensures that the futures trader realizes an almost immediate profit corresponding to the increase in the futures price. If the forward trader closed out the position by entering into a short contract at 1.3040, the forward trader would have contracted to buy 1 million pounds at 1.3000 in three months and sell 1 million pounds at 1.3040 in three months. This would lead to a \$4,000 profit—but in three months, not in one day. The forward trader's profit is the present value of \$4,000.

The forward trader can gain some consolation from the fact that gains and losses are treated symmetrically. If the forward/futures prices dropped to 1.2960 instead of rising to 1.3040, the futures trader would take a loss of \$4,000 while the forward trader would take a loss of only \$3,900. Also, over the three-month contract life, the total gain or loss from the futures contract and the forward contract would be the same.

Table 5.3 Comparison of Forward and Futures Contracts

Forward	Futures
Private contract between two parties	Traded on an exchange
Not standardized	Standardized contract
Usually one specified delivery date	Range of delivery dates
Settled at end of contract	Settled daily
Delivery or final cash settlement usually takes place	Contract is usually closed out prior to maturity
Some credit risk	Virtually no credit risk

Futures are cleared on an exchange clearing house. The clearing house stands between the two sides and is responsible for ensuring that required payments are made. The clearing house has a number of members. If a trader or broker is not a member of the exchange clearing house, it must arrange to clear trades through a member.

The clearing house requires initial margin and variation margin from its members. The variation margin for a day may be positive or negative and covers the gains and losses during the day. The initial margin is an extra amount held by the exchange that provides protection to the exchange in the event that the member defaults. In addition, the exchange clearing house requires default fund (also known as a guaranty fund) contributions from its members. This provides additional protection to the exchange. If a member defaults and its initial margin and default fund contributions are not enough to cover losses, the default fund contributions of other members can be used.

Exchange clearing house members require margin from brokers and other traders when they agree to clear their trades and brokers require margin from their clients. Typically the relationship between broker and client involves the posting of initial margin (greater than the initial margin applicable to a member). When the balance in the client's margin account (adjusted for daily gains and losses) falls below a maintenance margin level, the client is required to bring the balance back up to the initial margin level.

5.3.3 LIBOR and Its Replacements

Before moving on to discuss swaps and other derivatives, it is appropriate to discuss LIBOR (short for *London Interbank Offered Rate*). LIBOR rates were for 40 years compiled by asking a panel of global banks to provide quotes estimating the unsecured rates of interest at which they could borrow from other banks just prior to 11 A.M. (U.K. time). Five different currencies (the U.K. pound sterling, Swiss franc, euro, Japanese yen, and U.S. dollar) and several different borrowing periods (overnight, one week, one month, two months, three months, six months, and one year) were considered. The banks submitting quotes typically had good (AA) credit ratings. LIBOR rates were therefore considered to be estimates of unsecured borrowing rates for creditworthy banks.

LIBOR rates served as key reference rates for hundreds of trillions of dollars of transactions involving floating interest rates throughout the world. For example, the borrowing rate on a five-year loan in a particular situation might be specified as three-month LIBOR plus 30 basis points (i.e., three-month LIBOR plus 0.3%). The value of three-month LIBOR would then be noted at the beginning of each three-month period and interest based on this LIBOR rate would be paid by the borrower at the end of the period.

A problem with LIBOR was that there was not enough borrowing between banks for a bank's estimates to be based on market transactions. As a result, LIBOR submissions by banks involved a certain amount of judgment, and some banks were found guilty of manipulating their LIBOR quotes to increase their profits or for other reasons. As a

result, bank regulators insisted that the market move away from LIBOR to reference rates based on actual market transactions. In 2021, LIBOR was phased out as a reference rate for new transactions, although USD LIBOR continued to be quoted until mid-2023 for some maturities.

The regulators' plan was to base reference rates on overnight rates, which do trade actively. In the United States, the rate chosen was the secured overnight funding rate, SOFR; in the U.K., it was the sterling overnight index average, SONIA; in the eurozone, it was the euro short-term rate, ESTER; in Switzerland, it was the Swiss average rate overnight, SARON; in Japan, it was the Tokyo average overnight rate, TONAR.

Longer rates such as three-month rates, six-month rates, or one-year rates can be determined from overnight rates by compounding them daily. However there are two key differences between the use of LIBOR and the use of overnight rates for defining floating rates:

- LIBOR is known at the beginning of the period to which it applies. A rate calculated by compounding overnight rates is known only at the end of the period to which it applies. A borrower who agrees to pay three-month SOFR plus 50 basis points will therefore not know the rate for a particular three-month period until the end of that period.
- LIBOR incorporates some credit risk. The replacement overnight rates and the longer-term rates calculated from them are generally assumed to be risk-free rates. Indeed, they are referred to as RFRs (short for risk-free rates)

Other reference rates, which are similar to LIBOR but based on actual transactions have been suggested. These include the *Bloomberg Short-Term Yield Index* (BSBY), which estimates the average yield at which large banks access unsecured funds and *AMERIBOR*, created by the American Financial Exchange, which reflects the actual borrowing costs of thousands of community and regional banks in the United States. It remains to be seen whether these become more popular than the overnight rates favored by regulators.

5.3.4 Swaps

The first swap contracts were negotiated in the early 1980s. Since then, the market has seen phenomenal growth. Swaps now occupy a position of central importance in the over-the-counter derivatives market.

A swap is an agreement between two companies to exchange cash flows in the future. The agreement defines the dates when the cash flows are to be paid and the way in which they are to be calculated. Generally, the calculation of the cash flows involves the future values of interest rates, exchange rates, or other market variables.

A forward contract can be viewed as a simple example of a swap. Suppose it is March 1, 2024, and a company enters into a forward contract to buy 100 ounces of gold for \$1,300 per ounce in one year. The company can sell the gold in one year as

soon as it is received. The forward contract is therefore equivalent to a swap where the company agrees that on March 1, 2025, it will swap 100 times the spot price of gold for \$130,000.

Whereas a forward contract is equivalent to the exchange of cash flows on just one future date, swaps typically lead to cash flow exchanges taking place on several future dates. The most common swap is a “plain vanilla” interest rate swap where a fixed rate of interest is exchanged for a floating reference rate. Both interest rates are applied to the same notional principal. (The principal is “notional” because it is used only for the determination of interest exchanges. The principal itself is not exchanged.) A swap where Company A pays a fixed rate of interest of 3% and receives SOFR, with payments being exchanged every six months, is shown in Figure 5.3. (As explained earlier SOFR is an overnight rate in the United States. The six month rate is calculated by compounding daily rates and is not known at the end of the period to which it applies.) Suppose that in this contract the notional principal is \$100 million, and the swap lasts for three years. Table 5.4 shows the cash flows to Company A for one possible set of future interest rates. The swap is entered into on March 3, 2024. The SOFR floating interest rate for the first six months (known at the end of the six months) is assumed to be 2.2% per year or 1.1% per six months. (Note that all rates in Table 5.4 are assumed to be semiannually compounded.) As a result, the floating-rate cash flow received on September 3, 2024, is 0.011×100 or \$1.1 million. Similarly, SOFR for the second six months (known on

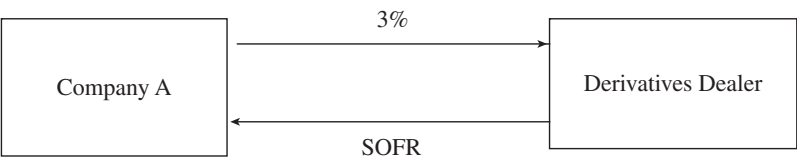


Figure 5.3 A Plain Vanilla Interest Rate Swap

Table 5.4 Example of Cash Flows (\$ millions) to Company A in Swap in Figure 5.3. Swap lasts three years and has a principal of \$100 million.

Date	Floating Cash Flow Received	Fixed Cash Flow Paid	Net Cash Flow
Mar. 3, 2024			
Sep. 3, 2024	+1.10	−1.50	−0.40
Mar. 3, 2025	+1.40	−1.50	−0.10
Sep. 3, 2025	+1.65	−1.50	+0.15
Mar. 3, 2026	+1.75	−1.50	+0.25
Sep. 3, 2026	+1.80	−1.50	+0.30
Mar. 3, 2027	+1.95	−1.50	+0.45

Cash flows do not take account of day count conventions, holiday calendars, and so on. Interest rates are semiannually compounded.

March 3, 2025) is assumed to be 2.8% per annum (or 1.4% per six months). This leads to a floating-rate cash flow of \$1.4 million; and so on. The fixed-rate cash flow paid is always \$1.5 million (3% of \$100 million applied to a six-month period).

Plain vanilla interest rate swaps are very popular because they can be used for many purposes. For example, the swap in Figure 5.3 could be used by Company A to transform borrowings at a floating rate of SOFR plus 1% to borrowings at a fixed rate of 4%. The combination of

1. Pay interest at SOFR plus 1% under loan agreement;
2. Receive SOFR under swap agreement; and
3. Pay 3% under the swap agreement

nets out to a fixed payment of 4%. It can also be used by Company A to transform an investment earning a fixed rate of 2.5% to an investment earning SOFR minus 0.5%. The combination of

1. Receive 2.5% on the investment;
2. Receive SOFR under swap agreement; and
3. Pay 3% under the swap agreement

nets out to a floating income at the rate of SOFR minus 0.5%.

Example 5.1

Suppose a bank has floating-rate deposits and five-year fixed-rate loans. As will be discussed in Chapter 14, this exposes the bank to significant risks. If rates rise, the deposits will be rolled over at high rates and the bank's net interest income will contract. The bank can hedge its risks by entering into the interest rate swap in Figure 5.3 (taking the role of Company A). The swap can be viewed as transforming the floating-rate deposits to fixed-rate deposits. Alternatively, it can be viewed as transforming the fixed-rate loans to floating-rate loans.

Table 5.5 shows quotes for U.S. dollar swaps that might be posted by the derivatives dealer in Figure 5.3. The first row shows that the bank is prepared to enter into a two-year swap where it pays a fixed rate of 2.55% and receives SOFR. It is also prepared to enter into a swap where it receives 2.58% and pays SOFR. The bid-ask spread in Table 5.5 is 3 or 4 basis points. The average of the bid and asked fixed rates is known as the *swap rate*. This is shown in the final column of the table.

The valuation of plain vanilla interest rate swaps is discussed in Appendix D at the end of this book.

Table 5.5 Swap Quotes Made by a Market Maker (percent per annum) for SOFR Swaps

Maturity (years)	Bid	Ask	Swap Rate
2	2.55	2.58	2.565
3	2.97	3.00	2.985
4	3.15	3.19	3.170
5	3.26	3.30	3.280
7	3.40	3.44	3.420
10	3.48	3.52	3.500

5.3.5 Options

Options are traded both on exchanges and in the over-the-counter market. There are two basic types of option. A *call option* gives the holder the right to buy the underlying asset by a certain date for a certain price. A *put option* gives the holder the right to sell the underlying asset by a certain date for a certain price. The price in the contract is known as the *exercise price* or *strike price*; the date in the contract is known as the *expiration date* or *maturity date*. *American options* can be exercised at any time up to the expiration date. *European options* can be exercised only on the expiration date itself.¹ Most of the options that are traded on exchanges are American. In the exchange-traded equity option market, one contract is usually an agreement to buy or sell 100 shares. European options are generally easier to analyze than American options, and some of the properties of an American option are frequently deduced from those of its European counterpart.

An *at-the-money option* is an option where the strike price equals the price of the underlying asset.² An *out-of-the-money option* is a call option where the strike price is above the price of the underlying asset or a put option where the strike price is below this price. An *in-the-money option* is a call option where the strike price is below the price of the underlying asset or a put option where the strike price is above this price.

It should be emphasized that an option gives the holder the right to do something. The holder does not have to exercise this right. By contrast, in a forward and futures contract, the holder is obligated to buy or sell the underlying asset. Note that whereas it costs nothing to enter into a forward or futures contract, there is a cost to acquiring an option. This cost is referred to as the *option premium*.

The largest exchange in the world for trading stock options is the Chicago Board Options Exchange (CBOE; www.cboe.com). Table 5.6 gives the mid-market prices of some of the American options trading on Intel (ticker INTC) at a particular time on January 24, 2022. The option strike prices are \$45, \$47.5, \$50, \$52.5, \$55, and \$57.5.

¹Note that the terms *American* and *European* do not refer to the location of the option or the exchange. For example, a few of the options trading on North American exchanges are European.

²Other definitions of “at-the-money” are on occasion used. For example, an at-the money option is sometimes defined as one where the present value of the strike price equals the asset price (the present value being determined by discounting the strike price from the end of the life of the option to the present). An at-the-money option is also sometimes defined as a call option with a delta of 0.5 or a put option with a delta of -0.5 . (See Chapter 15 for a definition of delta.)

Table 5.6 Prices of Options on Intel, January 24, 2022 (stock price = \$51.98)

Strike Price	Calls			Puts		
	March 22	May 22	July 22	March 22	May 22	July 22
45	7.77	8.47	8.87	1.03	1.96	2.35
47.5	5.77	6.67	7.17	1.55	2.64	3.15
50	4.05	5.10	5.65	2.32	3.60	4.15
52.5	2.64	3.77	4.37	3.42	4.77	5.35
55	1.61	2.73	3.30	4.90	6.25	6.80
57.5	0.93	1.92	2.48	6.72	7.90	8.47

The maturities are March 2022, May 2022, and July 2022. The March options have an expiration date of March 18, 2022; the May options have an expiration date of May 20, 2022; and the July options have an expiration date of July 15, 2022.³ Intel's stock price was \$51.98

Suppose an investor instructs a broker to buy one May call option contract on Intel with a strike price of \$52.50. The broker will relay these instructions to a trader at the CBOE. This trader will then find another trader who wants to sell one October call contract on Intel with a strike price of \$52.50, and a price will be agreed upon. We assume that the price is \$3.77, as indicated in Table 5.6. This is the price for an option to buy one share. In the United States, one stock option contract is a contract to buy or sell 100 shares. Therefore, the investor must arrange for \$377 to be remitted to the exchange through the broker. The exchange will then arrange for this amount to be passed on to the party on the other side of the transaction.

In our example, the investor has obtained at a cost of \$377 the right to buy 100 Intel shares for \$52.50 each. The party on the other side of the transaction has received \$377 and has agreed to sell 100 Intel shares for \$3.77 per share if the investor chooses to exercise the option. If the price of Intel does not rise above \$52.50 before May 20, 2022, the option is not exercised and the investor loses \$377. But if the Intel share price does well and the option is exercised when it is \$60, the investor is able to buy 100 shares at \$52.50 per share when they are worth \$60 per share. This leads to a gain of \$750, or \$373 when the initial cost of the options is taken into account.

If the investor is bearish on Intel, an alternative trade could be the purchase of one July put option contract with a strike price of \$50. From Table 5.6, we see that this would cost $100 \times \$4.15$ or \$415. The investor would obtain at a cost of \$415 the right to sell 100 Intel shares for \$50 per share prior to July 15, 2022. If the Intel share stays above \$50, the option is not exercised and the investor loses \$415. But if the investor is able to exercise when the stock price is \$40, the investor makes a gain of \$1,000 by buying 100 Intel shares at \$40 and selling them for \$50. The net profit after the cost of the options is taken into account is \$585.

³The exchange chooses the expiration date as the third Friday of the delivery month.

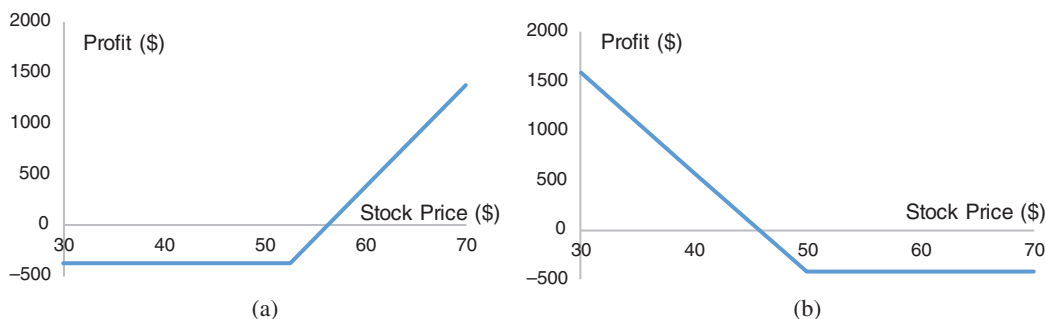


Figure 5.4 Net Profit from (a) Buying a Contract Consisting of 100 Intel October Call Options with a Strike Price of \$52.50 and (b) Buying a Contract Consisting of 100 Intel December Put Options with a Strike Price of \$50

The options trading on the CBOE are American. If we assume for simplicity that they are European so that they can be exercised only at maturity, the investor's profit as a function of the final stock price for the Intel options we have been considering is shown in Figure 5.4.

There are four types of trades in options markets:

1. Buying a call
2. Selling a call
3. Buying a put
4. Selling a put

Buyers are referred to as having *long positions*; sellers are referred to as having *short positions*. Selling an option is also known as *writing an option*.

When a trader purchases options for cash, there are no margin requirements because there are no circumstances under which the trade becomes a liability for the trader in the future. Options on stocks and stock indices that last longer than nine months can be bought on margin in the United States. The margin account works similarly to that for stocks except that 25% rather than 50% of the initial price can be borrowed.

When options are sold (i.e., written), there are potential future liabilities and margin must be posted. When a call option on a stock has been written, the initial and maintenance margin in the United States is the greater of

1. 100% of the value of the option plus 20% of the underlying share price less the amount, if any, by which the option is out-of-the-money.
2. 100% of the value of the option plus 10% of the share price.

When a put option is written, it is the greater of

1. 100% of the value of the option plus 20% of the underlying share price less the amount, if any, by which the option is out-of-the-money.
2. 100% of the value of the option plus 10% of the exercise price.

These margin requirements may be reduced if the trader has other positions in the stock. For example, if the trader has a fully covered position (where the trader has sold call

options on a certain number of shares and owns the same number of shares), there is no margin requirement on the short option position.

Options trade very actively in the over-the-counter market as well as on exchanges. The underlying assets include stocks, currencies, and stock indices. Indeed, the over-the-counter market for options is now larger than the exchange-traded market. Whereas exchange-traded options tend to be American, options trading in the over-the-counter market are frequently European. The advantage of the over-the-counter market is that maturity dates, strike prices, and contract sizes can be tailored to meet the precise needs of a client. They do not have to correspond to those specified by the exchange. The sizes of option trades in the over-the-counter are usually much greater than those on exchanges.

Valuation formulas and numerical procedures for options on a variety of assets are in Appendices E and F at the end of this book.

5.3.6 Interest Rate Options

Important interest rate options that trade in the over-the-counter market are *caps*, *floors*, and *swap options* (also known as *swaptions*). Whereas a swap exchanges a sequence of floating rates for a fixed rate, as indicated in Table 5.4, a cap, as its name implies, caps the floating rate. It is a series of call options on a floating rate. If the floating rate is greater than the strike rate (also known as cap rate), there is a payoff equal to the excess of the floating rate over the cap rate, applied to a predetermined notional for the appropriate accrual period; if the floating rate is less than the cap rate, there is no payoff. As in the case of swaps, payoffs are made at the end of the period covered by an interest rate.

A floor is similarly a series of put options on a floating rate. If the floating rate is less than the strike price (floor rate), there is a payoff equal to the difference between the two applied to the notional principal for the appropriate accrual period. Otherwise there is no payoff.

A swap option is an option to enter into a swap at some future time where the fixed rate is the strike rate. There are two types of swap options. One is the option to pay the strike rate and receive the floating rate; the other is the option to pay the floating rate and receive the strike rate. As in the case of a swap and cap or floor, a notional principal is specified.

5.4 Non-Traditional Derivatives

Whenever there are risks in the economy, financial engineers have attempted to devise derivatives to allow entities with exposures to manage their risks. Financial institutions typically act as intermediaries and arrange for the risks to be passed on to either (a) entities that have opposite exposures or (b) speculators who are willing to assume the risks. This section gives examples of derivatives that have been developed to meet specific needs.

5.4.1 *Weather Derivatives*

Many companies are in the position where their performance is liable to be adversely affected by the weather.⁴ It makes sense for these companies to consider hedging their weather risk in much the same way as they hedge foreign exchange or interest rate risks.

The first over-the-counter weather derivatives were introduced in 1997. To understand how they work, we explain two variables:

HDD: Heating degree days

CDD: Cooling degree days

A day's HDD is defined as

$$\text{HDD} = \max(0, 65 - A)$$

and a day's CDD is defined as

$$\text{CDD} = \max(0, A - 65)$$

where A is the average of the highest and lowest temperature during the day at a specified weather station, measured in degrees Fahrenheit. For example, if the maximum temperature during a day (midnight to midnight) is 68° Fahrenheit and the minimum temperature is 44° Fahrenheit, $A = 56$. The daily HDD is then 9 and the daily CDD is 0.

A typical over-the-counter product is a forward or option contract providing a pay-off dependent on the cumulative HDD or CDD during a month (that is, the total of the HDDs or CDDs for every day in the month). For example, a dealer could, in January 2024, sell a client a call option on the cumulative HDD during February 2025 at the Chicago O'Hare Airport weather station, with a strike price of 700 and a payment rate of \$10,000 per degree day. If the actual cumulative HDD is 820, the payoff is $\$10,000 \times (820 - 700) = \1.2 million. Often contracts include a payment cap. If the cap in our example is \$1.5 million, the client's position is equivalent to a long call option on cumulative HDD with a strike price of 700 and a short call option on cumulative HDD with a strike price of 850.

A day's HDD is a measure of the volume of energy required for heating during the day. A day's CDD is a measure of the volume of energy required for cooling during the day. Most weather derivatives contracts are entered into by energy producers and energy consumers. But retailers, supermarket chains, food and drink manufacturers, health service companies, agriculture companies, and companies in the leisure industry are also potential users of weather derivatives. The Weather Risk Management Association (www.wrma.org) has been formed to serve the interests of the weather risk management industry.

In September 1999, the Chicago Mercantile Exchange began trading weather futures and European options on weather futures. The contracts are on the cumulative HDD

⁴The U.S. Department of Energy has estimated that one-seventh of the U.S. economy is subject to weather risk.

and CDD for a month observed at a weather station. The contracts are settled in cash just after the end of the month once the HDD and CDD are known. One futures contract is \$20 times the cumulative HDD or CDD for the month. The CME now offers weather futures and options for many cities throughout the world. It also offers futures and options on hurricanes, frost, and snowfall.

5.4.2 Oil Derivatives

Crude oil is one of the most important commodities in the world. Ten-year fixed-price supply contracts have been commonplace in the over-the-counter market for many years. These are swaps where oil at a fixed price is exchanged for oil at a floating price.

There are many grades of crude oil, reflecting variations in gravity and sulfur content. Two important benchmarks for pricing are Brent crude oil (which is sourced from the North Sea) and West Texas Intermediate (WTI) crude oil. Crude oil is refined into products such as gasoline, heating oil, fuel oil, and kerosene.

In the over-the-counter market, virtually any derivative that is available on common stocks or stock indices is now available with oil as the underlying asset. Swaps, forward contracts, and options are popular. Contracts sometimes require settlement in cash and sometimes require settlement by physical delivery (i.e., by delivery of the oil).

Exchange-traded contracts on oil are also popular. The CME Group and the Intercontinental Exchange (ICE) trade a number of oil futures and futures options contracts. Some of the futures contracts are settled in cash; others are settled by physical delivery. For example, the Brent crude oil futures traded on ICE have a cash settlement option; the WTI oil futures traded by the CME Group require physical delivery. In both cases, the amount of oil underlying one contract is 1,000 barrels. The CME Group also trades popular contracts on two refined products: heating oil and gasoline. In both cases, one contract is for the delivery of 42,000 gallons.

5.4.3 Natural Gas Derivatives

The natural gas industry throughout the world went through a period of deregulation and the elimination of government monopolies in the 1980s and 1990s. The supplier of natural gas is now not necessarily the same company as the producer of the gas. Suppliers are faced with the problem of meeting daily demand.

A typical over-the-counter contract is for the delivery of a specified amount of natural gas at a roughly uniform rate over a one-month period. Forward contracts, options, and swaps are available in the over-the-counter market. The seller of gas is usually responsible for moving the gas through pipelines to the specified location.

The CME Group trades a contract for the delivery of 10,000 million British thermal units of natural gas. The contract, if not closed out, requires physical delivery to be made during the delivery month at a roughly uniform rate to a particular hub in Louisiana. ICE trades a similar contract in London.

Natural gas is a popular source of energy for heating buildings. It is also used to produce electricity, which in turn is used for air-conditioning. As a result, demand for natural gas is seasonal and dependent on the weather.

5.4.4 *Electricity Derivatives*

Electricity is an unusual commodity because it cannot easily be stored.⁵ The maximum supply of electricity in a region at any moment is determined by the maximum capacity of all the electricity-producing plants in the region. In the United States there are 140 regions known as *control areas*. Demand and supply are first matched within a control area, and any excess power is sold to other control areas. It is this excess power that constitutes the wholesale market for electricity. The ability of one control area to sell power to another control area depends on the transmission capacity of the lines between the two areas. Transmission from one area to another involves a transmission cost, charged by the owner of the line, and there are generally some energy transmission losses.

A major use of electricity is for air-conditioning systems. As a result, the demand for electricity, and therefore its price, is much greater in the summer months than in the winter months. The nonstorability of electricity causes occasional very large movements in the spot price. Heat waves have been known to increase the spot price by as much as 1,000% for short periods of time.

Like natural gas, electricity has been going through a period of deregulation and the elimination of government monopolies. This has been accompanied by the development of an electricity derivatives market. The CME Group now trades a futures contract on the price of electricity, and there is an active over-the-counter market in forward contracts, options, and swaps. A typical contract (exchange-traded or over-the-counter) allows one side to receive a specified number of megawatt hours for a specified price at a specified location during a particular month. In a 5×8 contract, power is received for five days a week (Monday to Friday) during the off-peak period (11 P.M. to 7 A.M.) for the specified month. In a 5×16 contract, power is received five days a week during the on-peak period (7 A.M. to 11 P.M.) for the specified month. In a 7×24 contract, power is received around the clock every day during the month. Option contracts have either daily exercise or monthly exercise. In the case of daily exercise, the option holder can choose on each day of the month (by giving one day's notice) to receive the specified amount of power at the specified strike price. When there is monthly exercise, a single decision on whether to receive power for the whole month at the specified strike price is made at the beginning of the month.

An interesting contract in electricity and natural gas markets is what is known as a *swing option* or *take-and-pay option*. In this contract, a minimum and maximum for the amount of power that must be purchased at a certain price by the option holder is

⁵Electricity producers with spare capacity sometimes use it to pump water to the top of their hydro-electric plants so that it can be used to produce electricity at a later time. This is the closest they can get to storing this commodity.

specified for each day during a month and for the month in total. The option holder can change (or swing) the rate at which the power is purchased during the month, but usually there is a limit on the total number of changes that can be made.

5.5 Exotic Options and Structured Products

Many different types of exotic options and structured products trade in the over-the-counter market. Although the amount of trading in them is small when compared with the trading in the plain vanilla derivatives, they are important to a bank because the profit on trades in exotic options and structured products tends to be much higher than on plain vanilla options or swaps.

Here are a few examples of exotic options:

Asian options: Whereas regular options provide a payoff based on the price of the underlying asset at the time of exercise, Asian options provide a payoff based on the average of the price of the underlying asset over some specified period. An example is an *average price call option* that provides a payoff in one year equal to $\max(\bar{S} - K, 0)$ where $\bar{S}$ is the average asset price during the year and K is the strike price.

Barrier options: These are options that come into existence or disappear when the price of the underlying asset reaches a certain barrier. For example, a knock-out call option with a strike price of \$30 and a barrier of \$20 is a regular European call option that ceases to exist if the asset price falls below \$20.

Basket options: These are options to buy or sell a portfolio of assets rather than options on a single asset.

Binary options: These are options that provide a fixed dollar payoff, or a certain amount of the underlying asset, if some condition is satisfied. An example is an option that provides a payoff in one year of \$1,000 if a stock price is greater than \$20.

Compound options: These are options on options. There are four types: a call on a call, a call on a put, a put on a call, and a put on a put. An example of a compound option is a European option to buy a European option on a stock currently worth \$15. The first option expires in one year and has a strike price of \$1. The second option expires in three years and has a strike price of \$20.

Lookback options: These are options that provide a payoff based on the maximum or minimum price of the underlying asset over some period. An example is an option that provides a payoff in one year equal to $S_T - S_{\min}$ where S_T is the asset price at the end of the year and $S_{\min}$ is the minimum asset price during the year.

Exotic options are sometimes more appropriate for hedging than plain vanilla options. As explained in Business Snapshot 5.3, Microsoft has used Asian options on a basket of some of its foreign currency hedging.

Structured products are products created by banks to meet the needs of investors or corporate treasurers. One example of a structured product is a principal protected note, where a bank offers an investor the opportunity to earn a certain percentage of the return

BUSINESS SNAPSHOT 5.3

Microsoft's Hedging

Microsoft actively manages its foreign exchange exposure. In some countries (e.g., Europe, Japan, and Australia), it bills in the local currency and converts its net revenue to U.S. dollars monthly. For these currencies, there is a clear exposure to exchange rate movements. In other countries (e.g., those in Latin America, Eastern Europe, and Southeast Asia), it bills in U.S. dollars. The latter appears to avoid any foreign exchange exposure—but it does not.

Suppose the U.S. dollar strengthens against the currency of a country in which Microsoft is billing in U.S. dollars. People in the country will find Microsoft's products more expensive because it takes more of the local currency to buy \$1. As a result, Microsoft is likely to find it necessary to reduce its (U.S. dollar) price in the country or face a decline in sales. Microsoft therefore has a foreign exchange exposure—both when it bills in U.S. dollars and when it bills in the local currency. (This shows that it is important for a company to consider the big picture when assessing its exposure.)

Microsoft sometimes uses options for hedging. Suppose it chooses a one-year time horizon. Microsoft recognizes that its exposure to an exchange rate (say, the Japanese yen–U.S. dollar exchange rate) is an exposure to the average of the exchange rates at the end of each month during the year. This is because approximately the same amount of Japanese yen is converted to U.S. dollars each month. Asian options, rather than regular options, are appropriate to hedge its exposure. What is more, Microsoft's total foreign exchange exposure is a weighted average of the exchange rates in all the countries in which it does business. This means that a basket option, where the option is on a portfolio of currencies, is an appropriate tool for hedging. An appropriate contract to negotiate with banks could therefore be an Asian basket option. The cost of this option is much less than a portfolio of put options, one for each month and each exchange rate (see Problem 5.23), but it gives the required protection.

Microsoft faces other financial risks. For example, it is exposed to interest rate risk on its bond portfolio. (When rates rise the portfolio loses money.) It also has two sorts of exposure to equity prices. It is exposed to the equity prices of the companies in which it invests. It is also exposed to its own equity price because it regularly repurchases its own shares as part of its stock awards program. It could use option strategies to hedge these risks.

BUSINESS SNAPSHOT 5.4

Procter and Gamble's Bizarre Deal

A particularly bizarre swap is the so-called 5/30 swap entered into by Bankers Trust (BT) and Procter and Gamble (P&G) on November 2, 1993. This was a five-year swap with semiannual payments. The notional principal was \$200 million. BT paid P&G 5.30% per annum. P&G paid BT the average 30-day CP (commercial paper) rate minus 75 basis points plus a spread. The average CP rate was calculated from observations on the 30-day commercial paper rate each day during the preceding accrual period.

The spread was zero for the first payment date (May 2, 1994). For the remaining nine payment dates, it was

$$\max \left[0, \frac{98.5 \left(\frac{5\text{-yr CMT}\%}{5.78\%} \right) - (30\text{-yr TSY Price})}{100} \right]$$

In this, five-year CMT is the constant maturity Treasury yield (that is, the yield on a five-year Treasury note, as reported by the U.S. Federal Reserve). The 30-year TSY price is the midpoint of the bid and ask cash bond prices for the 6.25% Treasury bond maturing in August 2023. Note that the spread calculated from the formula is a decimal interest rate. It is not measured in basis points. If the formula gives 0.1 and the average CP rate is 6%, the rate paid by P&G is 6% – 0.75% + 10% or 15.25%.

P&G was hoping that the spread would be zero and the deal would enable it to exchange fixed-rate funding at 5.30% for funding at 75 basis points less than the commercial paper rate. In fact, interest rates rose sharply in early 1994, bond prices fell, and the swap proved very, very expensive. (See Problem 5.38.)

provided by the S&P 500 with a guarantee that the return will not be negative. Another example of a (highly) structured product is the 5/30 transaction described in Business Snapshot 5.4.⁶ (In the case of this product, it is debatable whether Bankers Trust was meeting a client need or selling the client a product it did not need!)

⁶The details of this transaction are in the public domain because it later became the subject of litigation. See D. J. Smith, "Aggressive Corporate Finance: A Close Look at the Procter & Gamble–Bankers Trust Leveraged Swap," *Journal of Derivatives* 4, no. 4 (Summer 1997): 67–79.

BUSINESS SNAPSHOT 5.5

SocGen's Big Loss in 2008

Derivatives are very versatile instruments. They can be used for hedging, speculation, and arbitrage. One of the risks faced by a company that trades derivatives is that an employee who has a mandate to hedge or to look for arbitrage opportunities may become a speculator.

Jérôme Kerviel joined Société Générale (SocGen) in 2000 to work in the compliance area. In 2005, he was promoted and became a junior trader in the bank's Delta One products team. He traded equity indices such as the German DAX index, the French CAC 40, and the Euro Stoxx 50. His job was to look for arbitrage opportunities. These might arise if a futures contract on an equity index was trading for a different price on two different exchanges. They might also arise if equity index futures prices were not consistent with the prices of the shares constituting the index.

Kerviel used his knowledge of the bank's procedures to speculate while giving the appearance of arbitraging. He took big positions in equity indices and created fictitious trades to make it appear that he was hedged. In reality, he had large bets on the direction in which the indices would move. The size of his unhedged position grew over time to tens of billions of euros.

In January 2008, his unauthorized trading was uncovered by SocGen. Over a three-day period, the bank unwound his position for a loss of 4.9 billion euros. This was, at the time, the biggest loss created by fraudulent activity in the history of finance. (Later in the year, much bigger losses from Bernard Madoff's Ponzi scheme came to light.)

Rogue trader losses were not unknown at banks prior to 2008. For example, in the 1990s Nick Leeson, who worked at Barings Bank, had a similar mandate to Jérôme Kerviel's. His job was to arbitrage between Nikkei 225 futures quotes in Singapore and Osaka. Instead he found a way to make big bets on the direction of the Nikkei 225 using futures and options, losing \$1 billion and destroying the 200-year-old bank. In 2002, it was found that John Rusnak at Allied Irish Bank had lost \$700 million from unauthorized foreign exchange trading. In 2011, Kweku Adoboli, a member of UBS's Delta One team, lost \$2.3 billion by engaging in activities very similar to those of Jérôme Kerviel. The lesson from these losses is that it is important to define unambiguous risk limits for traders and then be very careful when monitoring what they do to make sure that the limits are adhered to.

5.6 Risk Management Challenges

Instruments such as futures, forwards, swaps, options, and structured products are versatile. They can be used for hedging, for speculation, and for arbitrage. (Hedging involves reducing risks; speculation involves taking risks; and arbitrage involves attempting to lock in a profit by simultaneously trading in two or more markets.) It is this very versatility that can cause problems. Sometimes traders who have a mandate to hedge risks or follow an arbitrage strategy become (consciously or unconsciously) speculators. The results can be disastrous. One example of this is provided by the activities of Jérôme Kerviel at Société Générale (see Business Snapshot 5.5).

To avoid the type of problems Société Générale encountered is an important risk management challenge. Both financial and nonfinancial corporations must set up controls to ensure that derivatives are being used for their intended purpose. Risk limits should be set and the activities of traders should be monitored daily to ensure that the risk limits are adhered to. We will be discussing this in later chapters.

Summary

This chapter has reviewed spot trades, forwards, futures, swaps, and options contracts. A forward or futures contract involves an obligation to buy or sell an asset at a certain time in the future for a certain price. A swap is an agreement to exchange cash flows in the future in amounts dependent on the values of one or more market variables. There are two types of options: calls and puts. A call option gives the holder the right to buy an asset by a certain date for a certain price. A put option gives the holder the right to sell an asset by a certain date for a certain price.

Forward, futures, and swap contracts have the effect of locking in the prices that will apply to future transactions. Options by contrast provide insurance. They ensure that the price applicable to a future transaction will not be worse than a certain level. Exotic options and structured products are tailored to the particular needs of corporate treasurers. For example, as shown in Business Snapshot 5.3, Asian basket options can allow a company to hedge its net exposure to several risks over a period of time. Derivatives now trade on a wide variety of variables. This chapter has reviewed those that provide payoffs dependent on the weather, oil, natural gas, and electricity. It has also discussed exotic options and structured products.

Further Reading

- Boyle, P., and F. Boyle. *Derivatives: The Tools That Changed Finance*. London: Risk Books, 2001.
- Flavell, R. *Swaps and Other Instruments*. 2nd ed. Chichester, UK: John Wiley & Sons, 2010.
- Geczy, C., B. A. Minton, and C. Schrand. "Why Firms Use Currency Derivatives." *Journal of Finance* 52, no. 4 (1997): 1323–1354.
- Litzenberger, R. H. "Swaps: Plain and Fanciful." *Journal of Finance* 47, no. 3 (1992): 831–850.

- Miller, M. H. "Financial Innovation: Achievements and Prospects." *Journal of Applied Corporate Finance* 4 (Winter 1992): 4–11.
- Warwick, B., F. J. Jones, and R. J. Teweles. *The Futures Games*. 3rd ed. New York: McGraw-Hill, 1998.

Practice Questions and Problems **(Answers at End of Book)**

- 5.1 What is the difference between a long forward position and a short forward position?
- 5.2 Explain carefully the difference between hedging, speculation, and arbitrage.
- 5.3 What is the difference between entering into a long forward contract when the forward price is \$50 and taking a long position in a call option with a strike price of \$50?
- 5.4 Explain carefully the difference between selling a call option and buying a put option.
- 5.5 An investor enters into a short forward contract to sell 100,000 British pounds for U.S. dollars at an exchange rate of 1.3000 U.S. dollars per pound. How much does the investor gain or lose if the exchange rate at the end of the contract is (a) 1.2900 and (b) 1.3200?
- 5.6 A trader enters into a short cotton futures contract when the futures price is 50 cents per pound. The contract is for the delivery of 50,000 pounds. How much does the trader gain or lose if the cotton price at the end of the contract is (a) 48.20 cents per pound and (b) 51.30 cents per pound?
- 5.7 Suppose you write a put contract with a strike price of \$40 and an expiration date in three months. The current stock price is \$41 and the contract is on 100 shares. What have you committed yourself to? How much could you gain or lose?
- 5.8 What is the difference between the over-the-counter market and the exchange-traded market? Which of the two markets do the following trade in: (a) a forward contract, (b) a futures contract, (c) an option, (d) a swap, and (e) an exotic option?
- 5.9 You would like to speculate on a rise in the price of a certain stock. The current stock price is \$29, and a three-month call with a strike of \$30 costs \$2.90. You have \$5,800 to invest. Identify two alternative strategies, one involving an investment in the stock and the other involving an investment in the option. What are the potential gains and losses from each?
- 5.10 Suppose that you own 5,000 shares worth \$25 each. How can put options be used to provide you with insurance against a decline in the value of your holding over the next four months?
- 5.11 When first issued, a stock provides funds for a company. Is the same true of a stock option? Discuss.
- 5.12 Suppose that a March call option to buy a share for \$50 costs \$2.50 and is held until March. Under what circumstances will the holder of the option make a profit? Under what circumstances will the option be exercised?

- 5.13 Suppose that a June put option to sell a share for \$60 costs \$4 and is held until June. Under what circumstances will the seller of the option (i.e., the party with the short position) make a profit? Under what circumstances will the option be exercised?
- 5.14 A company knows that it is due to receive a certain amount of a foreign currency in four months. What type of option contract is appropriate for hedging?
- 5.15 A United States company expects to have to pay 1 million Canadian dollars in six months. Explain how the exchange rate risk can be hedged using (a) a forward contract and (b) an option.
- 5.16 In the 1980s, Bankers Trust developed index currency option notes (ICONs). These are bonds in which the amount received by the holder at maturity varies with a foreign exchange rate. One example was its trade with the Long-Term Credit Bank of Japan. The ICON specified that if the yen–U.S. dollar exchange rate, S_T , is greater than 169 yen per dollar at maturity (in 1995), the holder of the bond receives \$1,000. If it is less than 169 yen per dollar, the amount received by the holder of the bond is

$$1,000 - \max \left[0, 1,000 \left(\frac{169}{S_T} - 1 \right) \right]$$

When the exchange rate is below 84.5, nothing is received by the holder at maturity. Show that this ICON is a combination of a regular bond and two options.

- 5.17 Suppose that USD–sterling spot and forward exchange rates are as follows:

Spot	1.3080
90-day forward	1.3056
180-day forward	1.3018

What opportunities are open to an arbitrageur in the following situations?

- (a) A 180-day European call option to buy £1 for \$1.27 costs 2 cents.
- (b) A 90-day European put option to sell £1 for \$1.34 costs 2 cents.
- 5.18 A company has money invested at 3% for five years. It wishes to use the swap quotes in Table 5.5 to convert its investment to a floating-rate investment. Explain how it can do this.
- 5.19 A company has borrowed money for five years at 5%. Explain how it can use the quotes in Table 5.5 to convert this to a floating-rate liability.
- 5.20 A company has a liability that costs the floating-rate plus 1%. Explain how it can use the quotes in Table 5.5 to convert this to a three-year fixed-rate liability.
- 5.21 A corn farmer argues: “I do not use futures contracts for hedging. My real risk is not the price of corn. It is that my whole crop gets wiped out by the weather.” Discuss this viewpoint. Should the farmer estimate his or her expected production of corn and hedge to try to lock in a price for expected production?

- 5.22 An airline executive has argued: “There is no point in our hedging the price of jet fuel. There is just as much chance that we will lose from doing this as that we will gain.” Discuss the executive’s viewpoint.
- 5.23 Why is the cost of an Asian basket put option to Microsoft considerably less than the cost of a portfolio of put options, one for each currency and each maturity (see Business Snapshot 5.3)?
- 5.24 “Oil, gas, and electricity prices tend to exhibit mean reversion.” What do you think is meant by this statement? Which energy source is likely to have the highest rate of mean reversion? Which is likely to have the lowest?
- 5.25 Does a knock-out barrier call option become more or less valuable as the frequency with which the barrier is observed is increased?
- 5.26 Suppose that each day during July the minimum temperature is 68° Fahrenheit and the maximum temperature is 82° Fahrenheit. What is the payoff from a call option on the cumulative CDD during July with a strike of 250 and a payment rate of \$5,000 per degree day?
- 5.27 Explain how a 5×8 option contract on electricity for May 2024 with daily exercise works. Explain how a 5×8 option contract on electricity for May 2024 with monthly exercise works. Which is worth more?
- 5.28 A U.S. investor writes five call option contracts (i.e., options to buy 500 shares). The option price is \$3.50, the strike price is \$60, and the stock price is \$57. What is the initial margin requirement?
- 5.29 A trader shorts 500 shares of a stock when the price is \$50. The initial margin is 160% and the maintenance margin is 130%. How much margin is required from the investor initially? How high does the price of the stock have to rise for there to be a margin call?
- 5.30 What is the difference between the margin required by an exchange from one of its members for a future contract and the margin required by a broker from one of its clients?

Further Questions

- 5.31 A company enters into a short futures contract to sell 5,000 bushels of wheat for 250 cents per bushel. The initial margin is \$3,000 and the maintenance margin is \$2,000. What price change would lead to a margin call? Under what circumstances could \$1,500 be withdrawn from the margin account?
- 5.32 A trader buys 200 shares of a stock on margin. The price of the stock is \$20 per share. The initial margin is 60% and the maintenance margin is 30%. How much money does the trader have to provide initially? For what share price is there a margin call?
- 5.33 The current price of a stock is \$94, and three-month European call options with a strike price of \$95 currently sell for \$4.70. An investor who feels that the price of the stock will increase is trying to decide between buying 100 shares and buying 2,000 call options (= 20 contracts). Both strategies involve an investment of \$9,400.

- What advice would you give? How high does the stock price have to rise for the option strategy to be more profitable?
- 5.34 A bond issued by Standard Oil worked as follows. The holder received no interest. At the bond's maturity the company promised to pay \$1,000 plus an additional amount based on the price of oil at that time. The additional amount was equal to the product of 170 and the excess (if any) of the price of a barrel of oil at maturity over \$25. The maximum additional amount paid was \$2,550 (which corresponds to a price of \$40 per barrel). Show that the bond is a combination of a regular bond, a long position in call options on oil with a strike price of \$25, and a short position in call options on oil with a strike price of \$40.
- 5.35 The price of gold is currently \$1,500 per ounce. The forward price for delivery in one year is \$1,700. An arbitrageur can borrow money at 5% per annum. What should the arbitrageur do? Assume that the cost of storing gold is zero and that gold provides no income.
- 5.36 A company's investments earn the floating-rate minus 0.5%. Explain how it can use the quotes in Table 5.5 to convert them to (a) three-, (b) five-, and (c) ten-year fixed-rate investments.
- 5.37 What position is equivalent to a long forward contract to buy an asset at K on a certain date and a long position in a European put option to sell it for K on that date?
- 5.38 Estimate the interest rate paid by P&G on the 5/30 swap in Business Snapshot 5.4 if (a) the CP rate is 6.5% and the Treasury yield curve is flat at 6% and (b) the CP rate is 7.5% and the Treasury yield curve is flat at 7% with semiannual compounding.

Chapter 6

The OTC Derivatives Market

The over-the-counter (OTC) market is a huge network of traders who work for financial institutions, large corporations, or fund managers. As indicated in Chapter 5, the OTC derivatives market, when measured in terms of the value of the underlying asset, is much larger than the exchange-traded derivatives market. A key advantage of the over-the-counter market is that the terms of a contract do not have to be those specified by an exchange. Market participants are free to negotiate any mutually attractive deal. Phone conversations in the over-the-counter market are usually taped. If there is a dispute over an agreement reached by phone, the tapes are replayed to resolve the issue. Trades in the over-the-counter market are typically much larger than trades in the exchange-traded market.

Before the 2007–2008 Global Financial Crisis (GFC), the OTC market was largely unregulated. Two market participants could enter into any trade they liked. They could agree to post collateral or not post collateral. They could agree to clear the trade directly with each other or use a third party. Also, they were under no obligation to disclose details of the trade to anyone else. As we will explain, this has now changed.

6.1 A Reference Point: Exchange-Traded Markets

Before discussing OTC markets, it is useful to review the operation of exchange-traded markets so that similarities and differences become apparent. Exchanges have been used

to trade financial products for many years. Some exchanges, such as the New York Stock Exchange (NYSE; www.nyse.com), focus on the trading of stocks. Others, such as the Chicago Board Options Exchange (CBOE; www.cboe.com) and CME group (CME; www.cmegroup.com), are concerned with the trading of derivatives such as futures and options.

The role of the exchange is to define the contracts that trade and organize trading so that market participants can be sure that the trades they agree to will be honored. Traditionally individuals have met at the exchange and agreed on the prices for trades, often by using an elaborate system of hand signals. This is known as the *open outcry system*. Exchanges have now largely replaced the open outcry system with *electronic trading*. This involves traders entering their desired trades at a keyboard and a computer being used to match buyers and sellers. Not everyone agrees that the shift to electronic trading is desirable. Electronic trading is less physically demanding than traditional floor trading. However, in some ways, it is also less exciting. Traders do not have the opportunity to attempt to predict short-term market trends from the behavior and body language of other traders.

Sometimes trading is facilitated with market makers. These are individuals or companies who are always prepared to quote both a *bid price* (price at which they are prepared to buy) and an *ask* or *offer price* (price at which they are prepared to sell). For example, at the request of a trader, a market maker might quote “bid 30.30, ask 30.40” for a particular share indicating a willingness to buy at \$30.30 and sell at \$30.40. At the time quotes are provided to the trader the market maker does not know whether the trader wishes to buy or sell. Typically, the exchange will specify an upper bound for the spread between a market maker’s bid and ask prices. The market maker earns its profit from this spread, but must manage its inventories carefully to limit its exposure to price changes.

The exchange clearing house is an adjunct to the exchange. It has a number of members. Brokers who are not members must channel their business through a member. The clearing house stands between the two sides to a transaction and guarantees the transaction. This is particularly important for futures and options. To guarantee the performance of a member for these transactions, a clearing house will require margin (i.e., collateral) from the member. The margin required is updated daily and sometimes several times during a day. Members will in turn require margin from the non-members (e.g., brokers) for which they are clearing trades. Brokers will require margin from their clients.

In determining the margin requirements from clearing house members, *netting* is usually applied. This means that the short positions the clearing house member is handling for clients are netted against the long positions for the purpose of calculating margin requirements. Suppose for example that the clearing house member has two clients, one with a long position of 20 in a particular futures contract, the other with a short position of 15 in the same futures contract. The margin would be calculated on the basis of a long position in 5 contracts. Margin is calculated for futures and short option positions so that the clearing house has a very high probability of not losing money in the case

of a member default. Clearing house members are required to contribute to a guaranty fund. This may be used by the clearing house in the event that a member defaults and the member's margin proves insufficient to cover losses.

The margining system has worked well. Traders at major exchanges have always had their contracts honored. Futures markets were tested on October 19, 1987, when the S&P 500 index declined by over 20%. The margin balances of S&P 500 futures traders with long positions were insufficient to cover their losses. Some refused to provide more margin when asked by their brokers to do so. As a result, some brokers went bankrupt because they were unable to meet margin calls on contracts entered into for clients. However, the clearing houses had sufficient funds to ensure that members with short futures were paid.

6.2 Clearing in OTC Derivatives Markets

We now describe how transactions are cleared in the OTC market. There are two main approaches: bilateral clearing and central clearing. In bilateral clearing, market participants clear transactions with each other. In central clearing, a third party, known as a *central counterparty* (CCP), stands between the two sides and plays a similar role to an exchange clearing house. The two alternatives are illustrated schematically in Figure 6.1. (This figure makes the simplifying assumption that there are only eight market participants and one CCP.)

6.2.1 Margin

We have introduced the term “margin” in connection with the collateral posted in the exchange-traded market. It is also used to describe the collateral (if any) posted in the OTC derivatives market.

Variation margin is the collateral posted to reflect the change in the value of a derivatives portfolio. Consider the situation where Party A is trading with Party B and the collateral agreement states that variation margin (with no threshold or minimum transfer amount) has to be posted by both sides.¹ This means that, if the value of outstanding transactions changes during a day so that they increase in value by $\$X$ to A (and therefore decrease in value by $\$X$ to B), B has to provide A with $\$X$ of acceptable collateral. The cumulative effect of variation margin is that, if outstanding derivatives have a value of $+\$V$ to A and $-\$V$ to B at a particular time, B should have posted a total of $\$V$ of collateral with A by that time.²

¹A threshold is a minimum value of the portfolio to one side before it can demand margin, and the minimum transfer amount is the minimum change in value necessary for a margin to have to be posted.

²In this context, note that if A buys an option from B for \$10,000, it must pay \$10,000 to B, but B must then return the \$10,000 to A as variation margin.

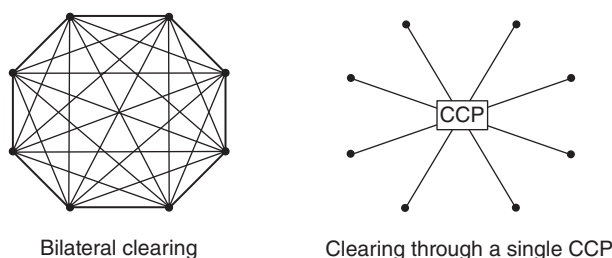


Figure 6.1 Bilateral and Central Clearing

Variation margin provides some protection against a counterparty default. It would provide total protection in an ideal world where (a) the counterparty never owes any variation margin at the time of default and (b) all outstanding positions can be replaced at mid-market prices as soon as the counterparty defaults. In practice, defaulting counterparties often stop posting collateral several days before they default, and the non-defaulting counterparty is usually subject to a bid-ask spread as it replaces transactions.³ To allow for adverse movements in the value of the portfolio during a period prior to default when no margin is being posted, OTC market participants sometimes require *initial margin* in addition to variation margin. Note that, in this context, adverse market movements are increases in the value of the portfolio to the non-defaulting party, not decreases. This is because increases in the value during a period when variation margin is not being posted lead to increases in replacement costs.⁴ Initial margin, which can change through time as the outstanding portfolio and relevant volatilities change, reflects the risk of a loss due to adverse market moves and the costs of replacing transactions.⁵

Most margin is cash, but the agreements in place may specify that securities can be posted instead of cash. The securities may be subject to a *haircut*. This means that the market value of the securities is reduced to determine their value for margin purposes. For example, a Treasury bond might be subject to a 10% haircut, indicating that, if its market value were \$100, it would cover only \$90 of a margin requirement.

Should cash margin earn interest? There is a difference between futures markets and OTC markets here. A futures exchange clearing house requires both initial margin and variation margin from members. Members earn interest on the initial margin. But they do not do so on variation margin because futures contracts are settled daily so that variation margin does not belong to the member posting it. OTC trades are not usually

³As explained later, the non-defaulting counterparty is able to claim from the defaulting party the costs related to the bid-ask spread that it would incur in replacing the transaction.

⁴It may seem strange that a market participant would be worried about the value of its transactions increasing. But suppose a transaction with a defaulting counterparty is hedged with another transaction entered into with another counterparty (as is often the case). The transaction with the other party can be expected to lose value without any compensating gain on the defaulted transaction.

⁵The non-defaulting party is allowed to keep all margin posted by the defaulting party up to the amount that can be legitimately claimed, as will be explained later.

settled daily and so interest is earned (paid) on the amount of variation margin provided (received).

6.2.2 Central Clearing

A CCP operates very much like an exchange clearing house. When two companies, A and B, agree to an over-the-counter derivatives transaction and decide to clear it centrally, they present it to a CCP. If the CCP accepts it (for a fee), it acts as an intermediary and enters into offsetting transactions with the two companies.

Suppose, for example, that the transaction is an interest rate swap where company A pays a fixed rate of 5% to company B on a principal of \$100 million for five years and company B pays a floating rate to company A on the same principal for the same period of time. Two separate transactions are created. Company A has a transaction with the CCP where it pays 5% and receives the floating rate on \$100 million. Company B has a transaction with the CCP where it pays the floating rate and receives 5% on \$100 million. The two companies no longer have credit exposure to each other. This is illustrated in Figure 6.2. If one or both parties to the transaction are not members of the CCP, they can clear the transaction through members.

Three large CCPs are

1. SwapClear (part of LCH Clearnet in London),
2. ClearPort (part of the CME Group in Chicago), and
3. ICE Clear Credit (part of the Intercontinental Exchange).

A CCP requires its members to provide initial margin and variation margin for the transactions being cleared. Typically, the initial margin is calculated so that there is a roughly 99% probability that it will cover market moves over five days. This protects the CCP from losses as it tries to close out or replace the positions of a defaulting member.

Consider the swap in Figure 6.2. Suppose for simplicity that it is the only transaction each side has with the CCP. The CCP might require an initial margin of \$0.5 million from each side. If, on the first day, interest rates fall so that the value of the swap to A

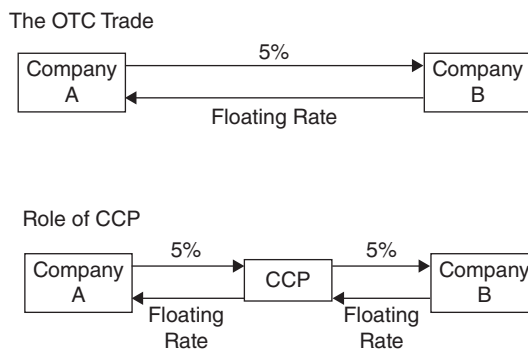


Figure 6.2 Role of CCP in OTC Markets

goes down by \$100,000, Party A would be required to pay a variation margin equal to this to the CCP, and the CCP would be required to pay the same amount to B. There could also be a change to the initial margin requirements determined by the CCP. If required margin is not paid by one of its members, the CCP closes out its transactions with that member. Cash and Treasury instruments are usually accepted as margin by CCPs. Typically the interest rate paid on cash balances is close to the overnight federal funds rate for U.S. dollars (and close to similar overnight rates for other currencies).

In practice, market participants are likely to have multiple transactions outstanding with the CCP at any given time. The initial margin required from a participant at any given time reflects the volatility of the value of its total position with the CCP. The role of a CCP in the OTC market is similar to the role of a clearing house in the exchange-traded market. The main difference is that transactions handled by the CCP are usually less standard than transactions in the exchange-traded market so that the calculation of margin requirements is more complicated.

The key advantage of clearing a transaction through a CCP is that OTC market participants do not need to worry about the creditworthiness of the counterparties they trade with. Credit risk is handled by the CCP using initial and variation margin.

A CCP requires its members to contribute to a default fund. (As mentioned, if one or both parties to a transaction are not members of the CCP, they can clear the transaction through members. They will then have to post margin with the members.) If a member fails to post margin when required, the member is in default and its positions are closed out. In closing out a member's positions, the CCP may incur a loss. A waterfall defines who bears the loss. The order in which the loss is funded is usually as follows:

1. The initial margin of the defaulting member
2. The default fund contribution of the member
3. The default fund contributions of other members
4. The equity of the CCP⁶

This is similar to the way losses in the event of a default are funded by an exchange clearing house.

6.2.3 *Bilateral Clearing*

In bilateral clearing, each pair of market participants enters into an agreement describing how all future transactions between them will be cleared. Typically this is an *ISDA master agreement*. (ISDA is short for International Swaps and Derivatives Association.) An annex to the agreement, known as the *credit support annex* (CSA), defines collateral arrangements. In particular, it defines what collateral (if any) has to be posted by each side, what assets are acceptable as collateral, what haircuts will be applied, and so on. The main

⁶In some cases, the non-defaulting members are required to provide additional default fund contributions when there is a default, with a cap on the amount of these additional contributions. (This is true of both exchange clearing houses and CCPs.)

body of the agreement defines what happens when one side defaults (e.g., by declaring bankruptcy, failing to make payments on the derivatives as they are due, or failing to post collateral when required). We will discuss this in more detail shortly.

6.2.4 Netting

Netting is a feature of ISDA master agreements and a feature of the agreements between CCPs and their members. It means that all transactions between two parties are considered to be a single transaction when (a) collateral requirements are being calculated and (b) early terminations occur because of a default. Netting reduces credit risk on derivatives portfolios because it means that the defaulting party cannot choose to default on transactions that have a negative value to it while keeping those that have a positive value.

Netting can also save initial margin. Suppose Party A has two transactions with a CCP that are not perfectly correlated. The initial margin for the portfolio is likely to be less than that for the two transactions separately.

6.2.5 Events of Default

Derivatives transactions are treated differently from other transactions in the event that a market participant fails to meet its obligations. For example, in ISDA master agreements there is an early termination provision that takes precedence over bankruptcy rules. This states that, if there is an “event of default,” the non-defaulting party has the right to terminate all transactions with the defaulting party after a short period of time has elapsed.⁷ Events of default include declarations of bankruptcy, failure to make payments as they are due, and failure to post collateral when required.⁸ Non-derivative contracts cannot always be terminated in this way. Another important difference between derivatives transactions and non-derivatives transactions is that in the case of derivatives transactions the non-defaulting party can take immediate possession of any collateral that has been posted by the defaulting party. It does not have to get a court order to allow it to do this.

If there is an event of default under an ISDA master agreement, the non-defaulting party calculates the mid-market value of outstanding transactions. It then adjusts this valuation in its favor by half the bid-ask spreads on the transactions for the purposes of calculating a settlement amount. This adjustment is compensation for the fact that it will have to trade with other dealers to replace the transactions and it will be subject to their bid-ask spreads when it does so. Suppose that one transaction has a mid-market value

⁷The non-defaulting party is not obliged to terminate transactions. When outstanding transactions have a negative value to them, counterparties sometimes consider that it is in their best interests not to terminate.

⁸Failure resolution mechanisms have been proposed where transactions are stayed (i.e., not terminated) for a period of time even if there is a bankruptcy filing, provided margin/collateral continues to be posted. These would allow the derivatives portfolios of bankrupt market participants to be unwound in an orderly way.

of \$20 million to the non-defaulting party and that the transaction is bid \$18 million, ask \$22 million. For the purposes of settlement, the transaction would be valued at \$22 million because this is what it would cost the non-defaulting party to replace the defaulting party's position in the transaction. If the non-defaulting party had the other side of the transaction so that its mid-market value were $-\$20$ million, it would be valued at $-\$18$ million for settlement purposes. In this case, the assumption is that a third party would be prepared to pay only \$18 million to take the defaulting party's position.

6.3 Post-Crisis Regulatory Changes

The OTC derivatives market was considered by many to have been partly responsible for the 2007–2008 GFC. When the G20 leaders met in Pittsburgh in September 2009 in the aftermath of the GFC, they wanted to reduce systemic risk by regulating the OTC market. The statement issued by the leaders after the meeting included the following paragraph:

All standardized OTC derivative contracts should be traded on exchanges or electronic trading platforms, where appropriate, and cleared through central counterparties by end-2012 at the latest. OTC derivative contracts should be reported to trade repositories. Non-centrally cleared contracts should be subject to higher capital requirements. We ask the FSB and its relevant members to assess regularly implementation and whether it is sufficient to improve transparency in the derivatives markets, mitigate systemic risk, and protect against market abuse.

The results of this were three major changes affecting OTC derivatives:

1. A requirement that all standardized OTC derivatives be cleared through CCPs. Standardized derivatives include plain vanilla interest rate swaps (which account for the majority of OTC derivatives traded) and default swaps on credit indices. The purpose of this requirement is to reduce the credit exposure derivatives dealers have to each other. Financial institutions are then less interconnected and there is less likely to be collapse of the financial system.
2. A requirement that standardized OTC derivatives be traded on electronic platforms. This is to improve transparency. The thinking is that, if there is an electronic platform for matching buyers and sellers, the prices at which products trade should be readily available to all market participants.⁹ The platforms are called *swap execution facilities* (SEFs) in the United States and *multilateral trading facilities* (MTFs) or *organized trading facilities* (OTFs) in Europe. In practice, standardized products once they have been traded on these platforms are passed automatically to a CCP.

⁹An issue here is that the type of electronic platform that is appropriate for swaps may not be the same as the one that is used by exchanges. Swaps are traded intermittently with large notional principals. Futures and options on an exchange trade continually and the size of trades is usually much smaller.

3. A requirement that all trades in the OTC market be reported to a central trade repository. This requirement provides regulators with important information on the risks being taken by participants in the OTC market. It is partly a response to the AIG situation. (AIG incurred huge losses in the OTC market, and regulators were unaware of them until the company sought a bailout in 2008.)

The first two of these requirements apply only to transactions between two financial institutions (or between a financial institution and a non-financial company that is considered to be systemically important because of the volume of its OTC derivatives trading). Derivatives dealers can therefore continue to trade with many of their non-financial corporate clients in the same way that they did pre-GFC.

About 25% of OTC transactions were cleared through CCPs pre-GFC and the remaining 75% were cleared bilaterally. As a result of the new rules, these percentages have flipped so that approximately 75% of OTC transactions are now cleared through CCPs while 25% are cleared bilaterally.

6.3.1 *Uncleared Trades*

Following another G20 meeting in 2011, the rules have been tightened for non-standard OTC derivatives. These are the derivatives that are not covered by the rules just mentioned. They are cleared bilaterally rather than centrally and are referred to as *uncleared trades*. Regulations, which have been implemented between 2016 and 2022, require uncleared trades between two financial institutions (or between a financial institution and a non-financial company that is considered to be systemically important) to be subject to rules on the margin that has to be posted. Previously, one of the attractions of bilateral clearing was that market participants were free to negotiate any credit support annex to their ISDA master agreements.

The rules state that both initial margin and variation margin must be posted for uncleared trades by both sides. Variation margin was fairly common in the OTC market pre-GFC (particularly in trades between derivatives dealers), but initial margin was rare. When entering into a transaction with a much less creditworthy counterparty, a derivatives dealer might insist on the counterparty posting initial margin. But the posting of initial margin by both sides was almost unheard of in the bilaterally cleared market.

Variation margin is usually transmitted directly from one counterparty to the other. Initial margin when posted by both sides cannot be handled in this way. If, for example, A transmitted \$1 million of initial margin to B and B transmitted \$1 million of initial margin to A, the initial margin would not serve the desired purpose because the transfers would cancel each other. For this reason the regulations require initial margin to be transmitted to a third party, where it is held in trust.

6.3.2 *Determination of Initial Margin: SIMM*

For the new rules on uncleared transactions to work, the two sides to an ISDA master agreement must agree on the variation margin and initial margin. The variation

margin requires agreement on the valuation of outstanding transactions, and procedures have been established for resolving any disagreements on this. The calculation of initial margin is more complicated than valuing the transactions and there is more scope for different models to give different results. As a result there have been attempts to develop an industry standard.

Initial margin is specified in the regulations for portfolios of uncleared transactions between two parties as the gain in value over 10 days that we are 99% certain will not be exceeded in stressed market conditions. Note that initial margin is the mirror image of VaR. When we are calculating VaR, we are determining extreme percentiles of the loss distribution, but when we are calculating initial margin we are determining extreme percentiles of the gain distribution. This is because exposure increases as the uncollateralized value of a portfolio increases.

The Basel Committee proposed a grid approach for calculating initial margin, which specified initial margin as a percentage of notional principal for different types of transactions. This was unpopular because it did not incorporate netting. If a market participant entered into a certain transaction on Day 1 and an almost offsetting transaction on Day 5, both with the same counterparty, the initial margin would double—even though the net exposure to the counterparty would be close to zero. ISDA proposed what is known as the Standard Initial Margin Model (SIMM) as a way of overcoming this. This model (see Appendix M) has now been approved by regulators.

6.4 Impact of the Changes

The new regulations have led to a world where more collateral is required for OTC derivatives transactions. Previously, most OTC transactions were cleared bilaterally and an initial margin was usually not required. Under the new regulations, most transactions are cleared through CCPs where both initial and variation margin are required from both sides. Furthermore, transactions that are cleared bilaterally between financial institutions require even more collateral than they would if they could be cleared through CCPs.

As discussed by Duffie and Zhu, there is one potential partial offset to the huge increase in collateral requirements mandated by the new rules.¹⁰ Under central clearing there is the potential for more netting. In Figure 6.1, under bilateral clearing a market participant has many different netting sets, one for each of the other market participants. Under central clearing, there is only one netting set. Bank A can, for example, net its transactions where Bank B is the counterparty with its transactions where Bank C is the counterparty, provided that all go through the same CCP.

Figure 6.1, however, is a simplification. It suggests that the choice is between a 100% bilateral world and a world where all transactions are cleared through a single CCP. The

¹⁰See D. Duffie and H. Zhu, “Does a Central Counterparty Reduce Counterparty Risk?” *Review of Asset Pricing Studies* 1 (2011): 74–95.

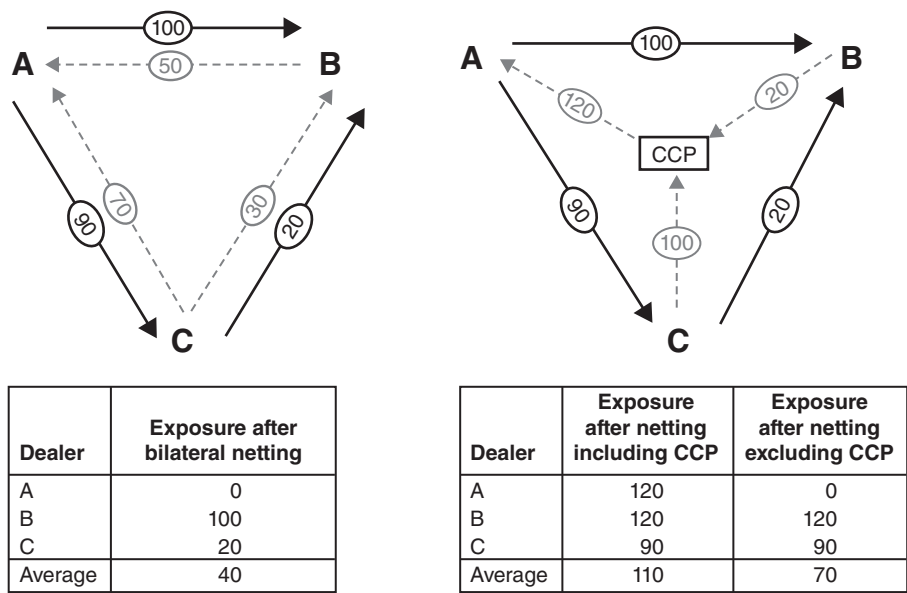


Figure 6.3 Example Where There Are Three Market Participants, One CCP, and Two Product Types
One product type (represented by dotted lines) can be cleared; the other (represented by solid lines) cannot.

reality is that (a) there will be a number of CCPs and they may not cooperate with each other to reduce initial margin requirements, and (b) some transactions will continue to be cleared bilaterally; so banks will face a situation that is a mixture of the two worlds depicted in Figure 6.1.

It is possible that the new rules requiring the use of CCPs will reduce rather than increase netting in some cases. This is illustrated by Figure 6.3, which shows the situation where there are three market participants and one CCP. The exposures represented by the dotted lines are standard transactions that can be cleared centrally. Those represented by the solid lines are nonstandard transactions that cannot be cleared centrally. For example, in B’s dealings with A, the nonstandard transactions are worth 100 to B and –100 to A; the standard transactions are worth +50 to A and –50 to B.

Without central clearing, the average exposure before collateral of the three parties is +40. With central clearing, the average exposure is 110 when the exposure to the CCP is included and 70 when it is not. Central clearing is likely to increase the collateral market participants have to post in this simple situation. This happens because without the central clearing rules standard transactions can be netted with nonstandard transactions, but with the central clearing rules this is no longer possible.

The overall effect of the changes has been an increase in margin requirements. Pre-GFC, relatively few OTC derivatives attracted initial margin. Now, the vast majority of OTC derivatives require initial margin. A related point is that financial institutions have more funds tied up in default fund contributions provided to CCPs.

BUSINESS SNAPSHOT 6.1

Rehypothecation

A practice in the management of collateral known as rehypothecation can cause problems. If Party A posts collateral with Party B and rehypothecation is permitted, Party B can use the same collateral to satisfy a demand for collateral from Party C; Party C can then use the collateral to satisfy a demand for collateral from Party D; and so on. In 2007, it was estimated that U.S. banks had more than \$4 trillion of collateral, but that this was created by using \$1 trillion of original collateral in conjunction with rehypothecation. Rehypothecation was particularly common in the United Kingdom, where title to collateral is transferred.

After Lehman declared bankruptcy in September 2008, clients (particularly European hedge fund clients) found it difficult to get a return of the collateral they had posted with Lehman because it had been rehypothecated. As a result of this experience, many market participants are more cautious than they used to be, and clauses in CSAs banning or limiting rehypothecation are now common.

6.4.1 Liquidity

An important consideration for all derivatives market participants is liquidity. Not only is the collateral posted at any given time a drain on liquidity, but banks have to keep a sufficient quantity of liquid assets on hand to ensure that they are able to meet any margin calls. (Margin calls from a CCP have to be met almost immediately.) As we will see in later chapters, regulators have recognized the importance of liquidity by developing two liquidity ratios that banks must adhere to.

6.4.2 Rehypothecation

What is known as “rehypothecation” was common in some jurisdictions (particularly the United Kingdom) before the GFC. (See Business Snapshot 6.1.) It involved a dealer using collateral posted with it by one counterparty to satisfy a collateral demand by another counterparty. It is estimated that pre-GFC about \$4 trillion of collateral was required in derivatives markets, but that because of rehypothecation only \$1 trillion of new collateral was posted.¹¹ In other words, each item of collateral was used on average four times. Rehypothecation is restricted under new rules developed since the crisis. These rules allow initial margin to be rehypothecated once, but only if certain

¹¹See M. Singh and J. Aitken, “The (Sizable) Role of Rehypothecation in the Shadow Banking System,” Working Paper, International Monetary Fund, 2010.

conditions are satisfied. Variation margin can be rehypothecated. But increasingly dealers themselves impose restrictions on rehypothecation because they do not want to be disadvantaged in the same way that some of Lehman's counterparties were (see Business Snapshot 6.1).

6.4.3 The Convergence of OTC and Exchange-Traded Markets

The developments we have been discussing are blurring the distinction between OTC derivatives and exchange-traded derivatives. Many OTC transactions are now traded on platforms similar to exchanges and cleared through organizations similar to exchange clearing houses. As time goes by, more OTC transactions are likely to be classified as "standard" so that the percentage of OTC transactions handled similarly to exchange-traded transactions will increase. What is more, even those OTC transactions between financial institutions that are cleared bilaterally may begin to look more like exchange-traded transactions. This is because initial margin has to be posted with a third party, and we can expect organizations (somewhat similar to exchange clearing houses) to be set up to facilitate this.

It is also the case that exchanges are increasingly trying to offer less standard products to institutional investors in an attempt to take business away from the OTC market. As a result, while OTC markets are moving in the direction of becoming more like exchange-traded markets, exchange-traded markets are moving in the opposite direction and becoming more like OTC markets. Many CCPs and exchanges have a common ownership and will find areas for cooperation on margin requirements and business practices. Whether a transaction is being cleared through an exchange or a CCP may not be important in the future because it will be handled in the same way by the same organization.

6.5 CCPs and Bankruptcy

The key objective of regulators is to reduce systemic risk. Some commentators have criticized the post-GFC derivatives regulations as replacing too-big-to-fail banks by too-big-to-fail CCPs. It certainly would be a disaster for the financial system if a major CCP such as LCH Clearnet's SwapClear or CME's ClearPort were to fail. In theory, as described in Hull (2012), it is possible to design the contract between CCPs and their members so that it is virtually impossible for a CCP to fail.¹² In practice, it is considered important that a CCP has "skin in the game." It is then motivated to take good decisions with respect to key issues such as whether a new member should be admitted, how initial margins should be set, and so on.

The main reason why it makes sense to replace too-big-to-fail banks by too-big-to-fail CCPs is that CCPs are much simpler organizations than banks. They are

¹²See J. Hull, "CCPs, Their Risks, and How They Can Be Reduced," *Journal of Derivatives* 20, no. 1 (Fall 2012): 26–29.

therefore much simpler to regulate than banks. In essence, regulators need ensure only that the CCP follows good practices in (a) choosing members, (b) valuing transactions, and (c) determining initial margins and default fund contributions. In the case of banks, a myriad of different, much more complex activities must be monitored. It is of course important for regulators to ensure that CCPs are not allowed to become more complex organizations by expanding outside their core activity of intermediating derivatives transactions.

Summary

Prior to the 2007–2008 Global Financial Crisis (GFC), the over-the-counter (OTC) derivatives market was largely unregulated. Two market participants could agree to any transaction they liked and then reach any agreement they liked on how the transaction would be cleared. They were also free to choose any arrangements they liked for the posting of collateral. This is no longer the case. The OTC derivatives market is now subject to a great deal of regulation throughout the world. The extent to which the OTC derivatives market should be blamed for the GFC is debatable, but post-GFC regulatory changes are having a big effect on this market.

Most standard OTC derivatives between two financial institutions must be cleared through central counterparties. These are very similar to exchanges. They require initial margin and variation margin to be posted by both sides. Nonstandard transactions between financial institutions will continue to be cleared bilaterally, but are subject to regulation on the collateral that must be posted. Specifically, transactions between financial institutions are subject to initial margin (segregated) and variation margin (transferred from one side to the other when the value of outstanding transactions changes).

What will the derivatives world look like in 15 or 20 years? Present trends suggest that there may be a convergence between OTC and exchange-traded markets, and the distinction between the two will become blurred. But there is of course no certainty that this trend will continue.

Further Reading

- Basel Committee on Banking Supervision and IOSCO. “Margin Requirements for Non-Centrally Cleared Derivatives,” September 2013.
- Duffie, D., and H. Zhu. “Does a Central Counterparty Reduce Counterparty Risk?” *Review of Asset Pricing Studies* 1 (2011): 74–95.
- Hull, J. “CCPs, Their Risks, and How They Can Be Reduced.” *Journal of Derivatives* 20, no. 1 (Fall 2012): 26–29.
- Hull, J. “The Changing Landscape for Derivatives.” *Journal of Financial Engineering* 1, no. 2 (2014).
- Hull, J. “OTC Derivatives and Central Clearing: Can All Transactions Be Cleared?” *Financial Stability Review* 14 (July 2010): 71–89.
- Singh, M., and J. Aitken. “The (Sizable) Role of Rehypotheication in the Shadow Banking System.” Working Paper, International Monetary Fund, 2010.

Practice Questions and Problems (Answers at End of Book)

- 6.1 What is the difference between using an ISDA master agreement and using a CCP for clearing OTC derivative transactions?
- 6.2 Explain the new regulations introduced since the 2007–2008 Global Financial Crisis that (a) require certain derivative transactions to be cleared using CCPs and (b) require extra collateral for some transactions cleared bilaterally.
- 6.3 Why might the regulations introduced since the 2007–2008 Global Financial Crisis create liquidity problems for some financial institutions?
- 6.4 What is meant by a haircut in a collateral agreement?
- 6.5 Explain the meaning of “event of default” and an “early termination” in an ISDA master agreement.
- 6.6 In Figure 6.3 where the CCP is used, suppose that half of the transactions between A and B that are represented by the solid line are moved to the CCP. What effect does this have on (a) the average exposure of the three parties including their exposures to the CCP and (b) the average exposure of the three parties excluding their exposures to the CCP?
- 6.7 What credit risks is a company taking when it becomes a member of a CCP and clears transactions through the CCP?
- 6.8 “Netting affects the collateral that has to be posted and the settlement in the event of an early termination.” Explain.
- 6.9 What is rehypothecation?
- 6.10 Why are CCPs easier to regulate than banks?
- 6.11 What is claimed by the non-defaulting party in an early termination under an ISDA master agreement?

Further Questions

- 6.12 “Some aspects of the new regulations can be expected to increase the amount of collateral posted for derivatives, and some can be expected to reduce it.” Explain this statement.
- 6.13 In Figure 6.3 where the CCP is used, suppose that an extra transaction between A and C that is worth 140 to A is cleared bilaterally. What effect does this have on the exposures?

Chapter 7

Securitization and the Global Financial Crisis

The Global Financial Crisis (GFC) started in the United States in 2007 and continued throughout 2008. The crisis spread rapidly to other countries and from financial markets to the real economy. It was the most serious financial crisis since the Great Depression of the 1930s. Some financial institutions failed. Others had to be bailed out. It is an episode in financial history that risk managers need to understand and learn from.

We saw in the previous chapter how the GFC led to regulations that transformed the over-the-counter (OTC) derivatives market. In later chapters, we will see how the GFC led to a major overhaul of the way banks are regulated and the way they manage risk. It therefore makes sense to examine the GFC at this stage in the book.

Derivatives such as futures, forwards, swaps, and options are one way in which risk can be transferred from one entity to another in the economy. Another is securitization. The securitization of mortgages in the United States played a major role in the GFC. In order to understand the GFC it is therefore necessary to understand how securitization works. This chapter therefore covers instruments such as asset-backed securities (ABSs) and collateralized debt obligations (CDOs).

7.1 The U.S. Housing Market

A natural starting point for a discussion of the Global Financial Crisis is the U.S. housing market. Figure 7.1 shows the S&P/Case-Shiller composite-10 index for house prices in the United States between 1987 and 2021. This tracks house prices for ten major metropolitan areas in the United States. In about the year 2000, house prices started to rise much faster than they had in the previous decade. The very low level of interest rates between 2002 and 2005 was an important contributory factor, but the bubble in house prices was largely fueled by mortgage lending practices.

The 2000 to 2006 period was characterized by a huge increase in what is termed subprime mortgage lending. Subprime mortgages are mortgages that are considered to be significantly more risky than average. Before 2000, most mortgages classified as subprime were second mortgages. After 2000, this changed as financial institutions became more comfortable with the notion of a subprime first mortgage.

7.1.1 The Relaxation of Lending Standards

Mortgage lenders in the United States started to relax their lending standards in about 2000. This made house purchases possible for many families that had previously been considered to be not sufficiently creditworthy to qualify for a mortgage. These families increased the demand for real estate and prices rose. To mortgage brokers and mortgage lenders, the combination of more lending and rising house prices was attractive. More lending meant bigger profits. Rising house prices meant that the lending was well covered by the underlying collateral. If the borrower defaulted, the resulting foreclosure would lead to little or no loss.

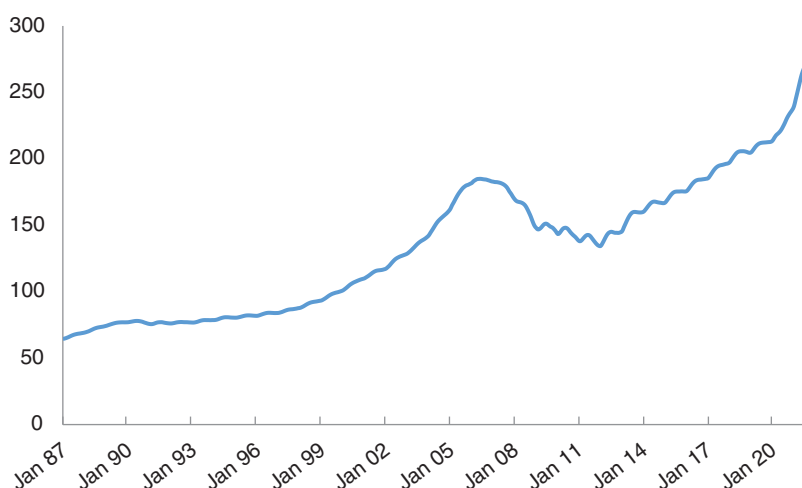


Figure 7.1 The S&P/Case-Shiller Composite-10 Index of U.S. Real Estate Prices (not seasonally adjusted), 1987 to November 2021. January 1, 2000 = 100.

How could mortgage brokers and mortgage lenders keep increasing their profits? Their problem was that, as house prices rose, it was more difficult for first-time buyers to afford a house. In order to continue to attract new entrants to the housing market, they had to find ways to relax their lending standards even more—and this is exactly what they did. The amount lent as a percentage of the house price increased. Adjustable rate mortgages (ARMs) were developed where there was a low teaser rate of interest that would last for two or three years and be followed by a rate that was liable to be much higher.¹ Lenders also became more cavalier in the way they reviewed mortgage applications. Indeed, the applicant's income and other information reported on the application were frequently not checked.

Why was the government not regulating the behavior of mortgage lenders? The answer is that the U.S. government had, since the 1990s, been trying to expand home ownership, and had been applying pressure to mortgage lenders to increase loans to low- and moderate-income households. Some state legislators (such as those in Ohio and Georgia) were concerned about what was going on and wanted to curtail predatory lending.² However, the courts decided that national standards should prevail.

A number of terms have been used to describe mortgage lending during the period leading up to the GFC. One is “liar loans,” because individuals applying for a mortgage, knowing that no checks would be carried out, sometimes chose to lie on the application form. Another term used to describe some borrowers is “NINJA” (no income, no job, no assets). Some analysts realized that the mortgages were risky, but pricing in the market for securities created from the mortgages suggests that the full extent of the risks and their potential impact on markets was not appreciated until well into 2007.

Mian and Sufi (2009) have carried out research confirming that there was a relaxation of the criteria used for mortgage lending.³ Their research defines “high denial zip codes” as zip codes where a high proportion of mortgage applicants had been turned down in 1996, and shows that mortgage origination grew particularly fast for these zip codes between 2000 to 2007. (Zip codes are postal codes in the United States defining the area in which a person lives.) Moreover, their research shows that lending criteria were relaxed progressively through time rather than all at once because originations in high-denial zip codes were an increasing function of time during the 2000 to 2007 period. Zimmerman (2007) provides some confirmation of this.⁴ He shows that subsequent default experience indicates that mortgages made in 2006 were of a lower quality than those made in 2005 and these were in turn of lower quality than mortgages made

¹A “2/28” ARM, for example, is an ARM where the rate is fixed for a two-year teaser-rate period and then floats for the remaining 28 years. If real estate prices increased, lenders expected the borrowers to prepay and take out a new mortgage at the end of the teaser rate period. However, prepayment penalties, often zero on prime mortgages, were quite high on subprime mortgages.

²Predatory lending describes the situation where a lender deceptively convinces borrowers to agree to unfair and abusive loan terms.

³See A. Mian and A. Sufi, “The Consequences of Mortgage Credit Expansion: Evidence from the US Mortgage Default Crisis,” *Quarterly Journal of Economics* 124, no. 4 (November 2009): 1449–1496.

⁴See T. Zimmerman, “The Great Subprime Meltdown,” *Journal of Structured Finance* (Fall 2007): 7–20.

in 2004. Standard & Poor's has estimated that subprime mortgage origination in 2006 alone totaled \$421 billion. AMP Capital Investors estimated that there was a total of \$1.4 trillion of subprime mortgages outstanding in July 2007.

7.1.2 *The Bubble Bursts*

The result of the relaxation of lending standards was an increase in the demand for houses and a bubble in house prices. Prices increased very fast during the 2000 to 2006 period. All bubbles burst eventually and this one was no exception. In the second half of 2006, house prices started to edge down. One reason was that, as house prices increased, demand for houses declined. Another was that some borrowers with teaser rates found that they could no longer afford their mortgages when the teaser rates ended. This led to foreclosures and an increase in the supply of houses for sale. The decline in house prices fed on itself. Individuals who had borrowed 100%, or close to 100%, of the cost of a house found that they had negative equity (i.e., the amount owed on the mortgage was greater than the value of the house). Some of these individuals chose to default. This led to more foreclosures, a further increase in the supply of houses for sale, and a further decline in house prices.

One of the features of the U.S. housing market is that mortgages are non-recourse in some states. This means that, when there is a default, the lender is able to take possession of the house, but other assets of the borrower are off-limits.⁵ Consequently, the borrower has a free American-style put option. He or she can at any time sell the house to the lender for the principal outstanding on the mortgage. (During the teaser-interest-rate period this principal typically increased, making the option more valuable.) Market participants realized belatedly how costly the put option could be. If the borrower had negative equity, the optimal decision was to exchange the house for the outstanding principal on the mortgage. The house was then sold, adding to the downward pressure on house prices.

It would be a mistake to assume that all mortgage defaulters were in the same position. Some were unable to meet mortgage payments and suffered greatly when they had to give up their homes. But many of the defaulters were speculators who bought multiple homes as rental properties and chose to exercise their put options. It was their tenants who suffered. There are also reports that some house owners (who were not speculators) were quite creative in extracting value from their put options. After handing the keys to their house to the lender, they turned around and bought (sometimes at a bargain price) another house that was in foreclosure. Imagine two people owning identical houses next to each other. Both have mortgages of \$250,000. Both houses are worth \$200,000 and in foreclosure can be expected to sell for \$170,000. What is the owners' optimal strategy? The answer is that each person should exercise the put option and buy the neighbor's house.

⁵In some other states, mortgages are not non-recourse but there is legislation making it difficult for lenders to take possession of other assets besides the house.

As foreclosures increased, the losses on mortgages also increased. Losses were high because houses in foreclosure were often surrounded by other houses that were also for sale. They were sometimes in poor condition. In addition, banks faced legal and other fees. In normal market conditions, a lender can expect to recover 75% of the amount owed in a foreclosure. In 2008 and 2009, recovery rates as low as 25% were experienced in some areas.

The United States was not alone in having declining real estate prices. Prices declined in many other countries as well. Real estate in the United Kingdom was particularly badly affected. As Figure 7.1 indicates, average house prices eventually recovered and were well above the previous peak by 2021.

7.2 Securitization

The originators of mortgages did not in many cases keep the mortgages themselves. They sold portfolios of mortgages to companies that created products for investors from them. This process is known as *securitization*. Securitization has been an important and useful tool for transferring risk in financial markets for many years. It underlies the originate-to-distribute model that was widely used by banks prior to 2007 and is discussed in Chapter 2.

Securitization played a part in the creation of the housing bubble. The behavior of mortgage originators was influenced by their knowledge that mortgages would be securitized.⁶ When considering new mortgage applications, the question was not: “Is this a credit we want to assume?” Instead it was: “Is this a mortgage we can make money from by selling it to someone else?”

When mortgages were securitized, the only information received about the mortgages by the buyers of the products that were created from them was the loan-to-value ratio (i.e., the ratio of the size of the loan to the assessed value of the house) and the borrower’s FICO (credit) score.⁷ The reason why lenders did not check information on things such as the applicant’s income, the number of years the applicant had lived at his or her current address, and so on, was that this information was considered irrelevant. The most important thing for the lender was whether the mortgage could be sold to others—and this depended primarily on the loan-to-value ratio and the applicant’s FICO score.

It is interesting to note in passing that both the loan-to-value ratio and the FICO score were of dubious quality. The property assessors who determined the value of a house at the time of a mortgage application sometimes inflated valuations because they

⁶Research by Keys et al. shows that there was a link between securitization and the lax screening of mortgages. See B. J. Keys, T. Mukherjee, A. Seru, and V. Vig, “Did Securitization Lead to Lax Screening? Evidence from Subprime Loans,” *Quarterly Journal of Economics* 125, no. 1 (February 2010): 307–362.

⁷FICO is a credit score developed by the Fair Isaac Corporation and is widely used in the United States. It ranges from 300 to 850.

knew that the lender wanted a low loan-to-value ratio. Potential borrowers were sometimes counseled to take certain actions that would improve their FICO scores.⁸

We now consider the products that were created from the mortgages and sold in the market.

7.2.1 *Asset-Backed Securities*

An *asset-backed security* (ABS) is a security created from the cash flows of financial assets such as loans, bonds, credit card receivables, mortgages, auto loans, and aircraft leases. Sometimes, cash flow streams such as royalties from the future sales of a piece of music are even used. The way the security works is illustrated in Figure 7.2. A portfolio of assets (such as subprime mortgages) is sold by the originators of the assets to a special purpose vehicle (SPV) and the cash flows from the assets are allocated to tranches. In Figure 7.2, there are three tranches. (This is a simplification. In reality there are usually many more than three tranches created.) These are the senior tranche, the mezzanine tranche, and the equity tranche. The portfolio has a principal of \$100 million. This is divided as follows: \$75 million to the senior tranche, \$20 million to the mezzanine tranche, and \$5 million to the equity tranche. The senior tranche is promised a return of 6%, the mezzanine tranche is promised a return of 10%, and the equity tranche is promised a return of 30%.

It sounds as though the equity tranche has the best deal, but this is not necessarily the case. The equity tranche is much less likely to realize its return than the other two tranches. Cash flows are allocated to tranches by specifying what is known as a waterfall.

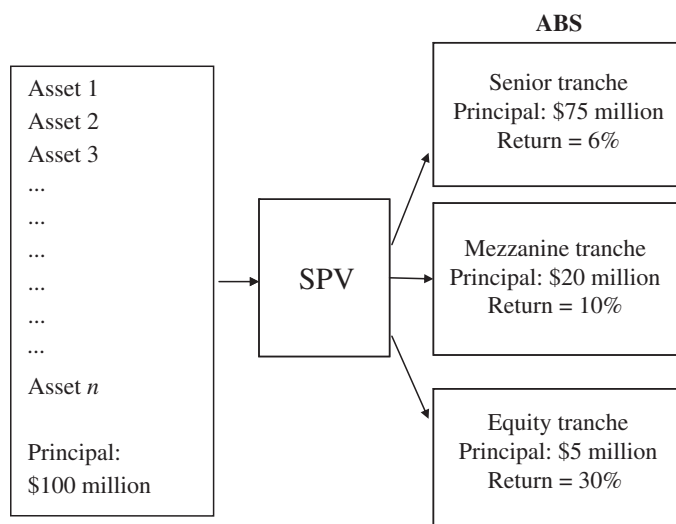


Figure 7.2 Creation of an Asset-Backed Security from a Portfolio of Assets (simplified)

⁸One such action might be to make regular payments on a credit card for a few months.

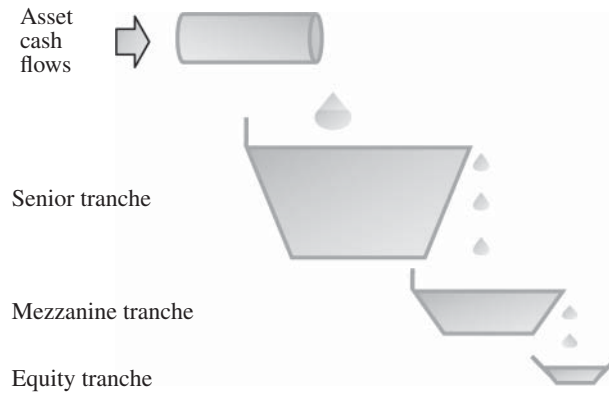


Figure 7.3 The Waterfall in an Asset-Backed Security

An approximation to the way a waterfall works is in Figure 7.3. There is a separate waterfall for interest and principal cash flows. Interest cash flows from the assets are allocated to the senior tranche until the senior tranche has received its promised return on its outstanding principal. Assuming that the promised return to the senior tranche can be made, cash flows are then allocated to the mezzanine tranche. If the promised return to the mezzanine tranche on its outstanding principal can be made and interest cash flows are left over, they are allocated to the equity tranche. Principal cash flows are used first to repay the principal of the senior tranche, then the mezzanine tranche, and finally the equity tranche.⁹

The structure in Figure 7.2 typically lasts several years. The extent to which the tranches get their principal back depends on losses on the underlying assets. The first 5% of losses are borne by the principal of the equity tranche. If losses exceed 5%, the equity tranche loses all its principal and some losses are borne by the principal of the mezzanine tranche. If losses exceed 25%, the mezzanine tranche loses all its principal and some losses are borne by the principal of the senior tranche.

There are therefore two ways of looking at an ABS. One is with reference to the waterfall in Figure 7.3. Cash flows go first to the senior tranche, then to the mezzanine tranche, and then to the equity tranche. The other is in terms of losses. Losses of principal are first borne by the equity tranche, then by the mezzanine tranche, and then by the senior tranche.

The ABS is designed so that the senior tranche is rated AAA. The mezzanine tranche is typically rated BBB. The equity tranche is typically unrated. Unlike the ratings assigned to bonds, the ratings assigned to the tranches of an ABS are what might be termed “negotiated ratings.” The objective of the creator of the ABS is to make the senior tranche as big as possible without losing its AAA credit rating. (This maximizes the profitability

⁹The priority rule described here is a simplification. The precise waterfall rules are somewhat more complicated and outlined in a legal document several hundred pages long.

of the structure.) The ABS creator examines information published by rating agencies on how tranches are rated and may present several structures to rating agencies for a preliminary evaluation before choosing the final one. The creator of the ABS expects to make a profit because the weighted average return on the assets in the underlying portfolio is greater than the weighted average return offered to the tranches.

A particular type of ABS is a *collateralized debt obligation* (CDO). This is an ABS where the underlying assets are fixed-income securities. The procedures used by the market to value a CDO are outlined in Appendix L.

7.2.2 ABS CDOs

Finding investors to buy the senior AAA-rated tranches created from subprime mortgages was not difficult. Equity tranches were typically retained by the originator of the mortgages or sold to a hedge fund. Finding investors for the mezzanine tranches was more difficult. This led financial engineers to be creative (arguably too creative). Financial engineers created an ABS from the mezzanine tranches of ABSs that were created from subprime mortgages. This is known as an *ABS CDO* or *Mezz ABS CDO* and is illustrated in Figure 7.4. (Like the ABS in Figure 7.3, this is simplified.) The senior tranche of the ABS CDO is rated AAA. This means that the total of the AAA-rated instruments created in the example that is considered here is 90% (75% plus 75% of 20%) of the principal of the underlying mortgage portfolios. This seems high but, if the securitization were carried further with an ABS being created from the mezzanine tranches of ABS CDOs (and this did happen), the percentage would be pushed even higher.

In the example in Figure 7.4, the AAA-rated tranche of the ABS would probably be downgraded in the second half of 2007. However, it would receive the promised return if losses on the underlying mortgage portfolios were less than 25% because all losses of principal would then be absorbed by the more junior tranches. The AAA-rated tranche of the ABS CDO in Figure 7.4 is much more risky. It will get paid the promised return only if losses on the underlying portfolios are 10% or less. If losses are more than 10%, the

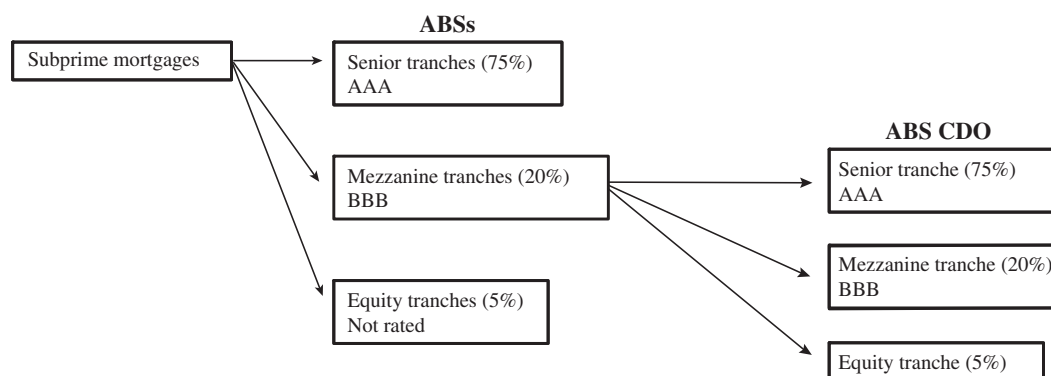


Figure 7.4 Creation of ABSs and an ABS CDO from Subprime Mortgages (simplified)

Table 7.1 Losses to Tranches in Figure 7.4

Losses to Subprime Portfolios	Losses to Mezzanine Tranche of ABS	Losses to Equity Tranche of ABS CDO	Losses to Mezzanine Tranche of ABS CDO	Losses to Senior Tranche of ABS CDO
10%	25%	100%	100%	0%
15%	50%	100%	100%	33%
20%	75%	100%	100%	67%
25%	100%	100%	100%	100%

equity tranche of the ABS is wiped out. The mezzanine tranche of the ABS bears more than 5% of the losses. This in turn means that the ABS CDO loses more than (5/20) or 25% of its principal so that the equity and mezzanine tranches of the ABS CDO are wiped out, and losses are incurred by the ABS CDO's senior tranche.

Consider, for example, the situation where losses are 20% on the underlying portfolios so that the mezzanine tranche of the ABS absorbs 15% of the losses. In this case, losses on the mezzanine tranches of the ABS CDO are 15/20 or 75% of their principal. The first 25% is absorbed by the equity and mezzanine tranches of the ABS CDO. The senior tranche of the ABS CDO therefore loses 50/75 or 67% of its value. These and other results are summarized in Table 7.1.

Many banks have lost money investing in the senior tranches of ABS CDOs. The investments typically promised a return quite a bit higher than the bank's funding cost. Because they were rated AAA, the capital requirements were minimal. Merrill Lynch is an example of a bank that lost a great deal of money from investments in ABS CDOs. In July 2008, Merrill Lynch agreed to sell senior tranches of ABS CDOs that had previously been rated AAA and had a principal of \$30.6 billion to Lone Star Funds for 22 cents on the dollar.¹⁰

7.2.3 CDOs and ABS CDOs in Practice

Figures 7.2 and 7.4 illustrate the nature of the securitizations that were done. In practice, many more tranches were created than those shown in Figures 7.2 and 7.4, and many of the tranches were thinner (i.e., corresponded to a narrower range of losses). Figure 7.5 shows a more realistic example of the structures that were created. This is adapted from an illustration by Gorton, which was taken from an article by UBS.¹¹

In Figure 7.5, two ABS CDOs are created. One (referred to as a Mezz ABS CDO) is created from the BBB-rated tranches of ABSs (similarly to the ABS CDO in Figure 7.4); the other (referred to as a high-grade ABS CDO) is from the AAA, AA, and A tranches of ABSs. The figure shows a third level of securitization based on the A and AA tranches of

¹⁰In fact the deal was worse than it sounds for Merrill Lynch because Merrill Lynch agreed to finance 75% of the purchase price. If the value of the tranches fell below 16.5 cents on the dollar, Merrill Lynch might have found itself owning the assets again.

¹¹G. Gorton, "The Subprime Panic," *European Financial Management* 15, no. 1 (2008): 10–46.

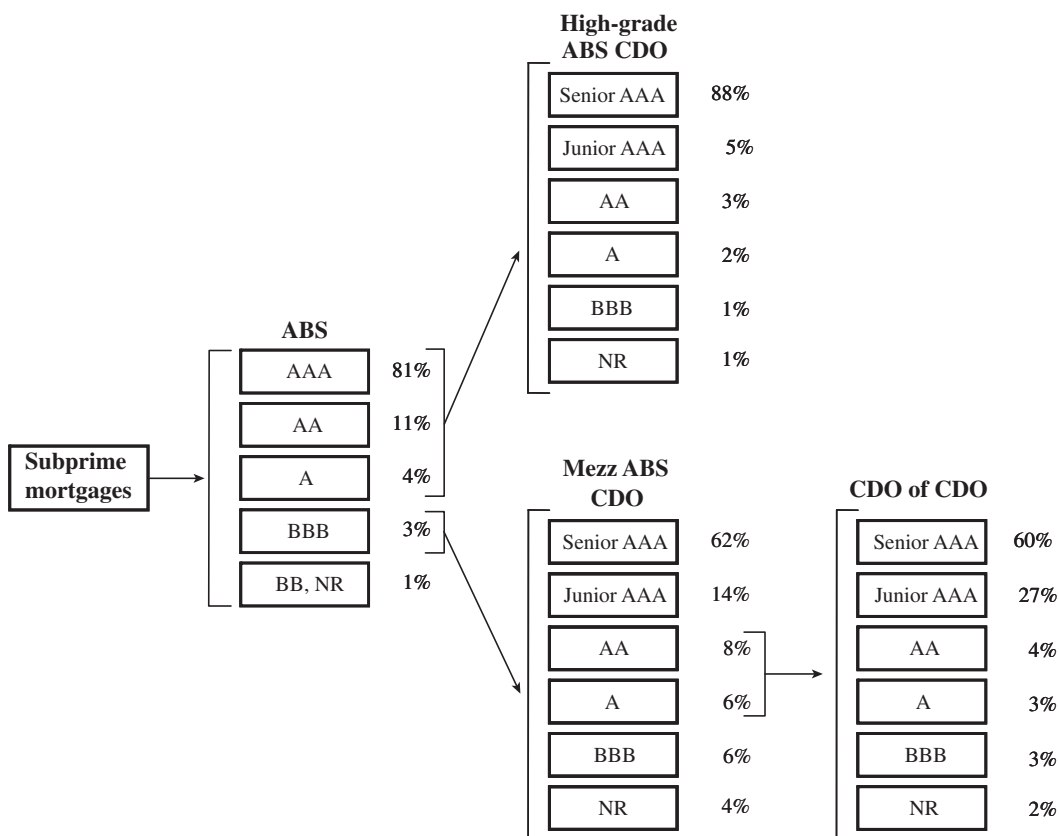


Figure 7.5 More Realistic Example of Subprime Securitizations with ABS, ABS CDOs, and a CDO of CDO Being Created

the Mezz ABS CDO. There was typically a small amount of *overcollateralization* with the face value of the mortgages being greater (by 1% or 2%) than the total face value of the ABS tranches. This created a cushion for investors, but by carrying out a similar analysis to that in Table 7.1 it is not difficult to see that investors in many of the tranches created would lose principal in situations where losses on the underlying subprime mortgage portfolios were moderately high.

The risks in the AAA-rated tranches of ABSs and ABS CDOs were higher than either investors or rating agencies realized. One of the reasons for this involves correlation. The values of the tranches of ABSs depend on the default correlation of the underlying mortgages. The tranches of ABS CDOs are even more heavily dependent on these default correlations. If mortgages exhibit a fairly low default correlation (as they do in normal times), there is very little chance of a high overall default rate and the AAA-rated tranches of both ABSs and ABS CDOs are safe. But, many analysts overlooked the fact that correlations always increase in stressed market conditions. In 2005 to 2006, the models used by investors and rating agencies assumed correlations that were too low for the upheavals in the U.S. housing market that were considered likely by many observers. As explained in Business Snapshot 7.1, another mistake made by analysts was

BUSINESS SNAPSHOT 7.1**All BBBs Are Not the Same**

Analysts tended to assume that the mezzanine tranche of an ABS, when rated BBB, can be considered to be identical to a BBB bond for the purposes of evaluating a CDO created from the mezzanine tranches. This is not a good assumption. The rating agency models attempted to ensure that the BBB tranche of an ABS had the same probability of loss, or the same expected loss, as a BBB bond. But the probability distribution of the loss is very different. Because the BBB tranches of ABSs were thin, it is much more likely that an investor in the BBB tranche of an ABS will lose everything than that this will happen for an investor in a BBB-rated bond. (This is sometimes referred to as “cliff risk.”) This means that the risk characteristics of ABS CDO tranches created from the BBB tranches of ABSs are quite different from the risk characteristics of similar tranches created from BBB bonds.

One lesson from this is that it is dangerous to interpret ratings for tranches of an ABS—or any other structured product—in the same way that ratings for bonds are interpreted. For similarly rated bonds and structured products, the probability distributions of losses are markedly different.

to assume that the BBB-rated tranches of an ABS were equivalent in risk to BBB-rated bonds. There are important differences between the two and these differences can have a big effect on the valuation of the tranches of ABS CDOs.

7.3 The Losses

The defaults on mortgages in the United States had a number of consequences. Financial institutions and other investors who had bought the tranches of ABSs and ABS CDOs lost money. Losses were also incurred by some mortgage originators because they had provided guarantees as to the quality of the mortgages that were securitized and because they faced lawsuits over their lending practices.

As often happens when losses are experienced in one segment of the debt market, there was a “flight to quality.” Investors became reluctant to take any credit risk and preferred to buy Treasury instruments and similarly safe investments. Credit spreads (the extra return required for taking credit risks) increased sharply. It was difficult for many non-financial companies to obtain loans from banks. Indeed, banks became reluctant to lend to each other at all and interbank lending rates increased sharply.

The tranches of ABSs and ABS CDOs were downgraded by rating agencies in the second half of 2007. The market for these tranches became very illiquid. Investors realized that they did not understand the tranches as well as they had previously thought and

that they had placed too much reliance on credit ratings. This emphasizes the importance of transparency in financial markets. The products created during the period leading up to the GFC were very complicated.¹² Investors did not worry about this until problems emerged. They then found that the liquidity of the market was such that they could only trade at fire-sale prices.

Banks such as Citigroup, UBS, and Merrill Lynch suffered huge losses. There were many government bailouts of financial institutions. Lehman Brothers was allowed to fail. The world experienced the worst recession since the 1930s. Unemployment increased. Even people in remote parts of the world that had no connection with U.S. financial institutions were affected.

Banks are now paying a price for the GFC. As we shall see in later chapters, they are required to keep more capital than before. They are also required to maintain certain liquidity ratios. Legislation such as Dodd–Frank in the United States increases the oversight of financial institutions and restricts their activities in areas such as proprietary trading.

7.4 What Went Wrong?

“Irrational exuberance” is a phrase coined by Alan Greenspan, Chairman of the Federal Reserve Board, to describe the behavior of investors during the bull market of the 1990s. It can also be applied to the period leading up to the GFC. Mortgage lenders, the investors in tranches of ABSs and ABS CDOs that were created from residential mortgages, and the companies that sold protection on the tranches assumed that the U.S. house prices would continue to increase—or at least not decrease. There might be declines in one or two areas, but the possibility of the widespread decline shown in Figure 7.1 was a scenario not considered by most people.

Many factors contributed to the crisis that started in 2007. Mortgage originators used lax lending standards. Products were developed to enable mortgage originators to profitably transfer credit risk to investors. Rating agencies moved from their traditional business of rating bonds, where they had a great deal of experience, to rating structured products, which were relatively new and for which there were relatively little historical data. The products bought by investors were complex and in many instances investors and rating agencies had inaccurate or incomplete information about the quality of the underlying assets. Investors in the structured products that were created thought they had found a money machine and chose to rely on rating agencies rather than forming their own opinions about the underlying risks. The return promised on the structured products rated AAA was high compared with the returns promised on bonds rated AAA.

¹²Some of the products that were created were even more complicated than indicated by the description in Section 7.2. For example, sometimes ABS CDO tranches were included in the portfolios used to create other ABS CDOs.

7.4.1 *Regulatory Arbitrage*

Many of the mortgages were originated by banks and it was banks that were the main investors in the tranches that were created from the mortgages. Why would banks choose to securitize mortgages and then buy the securitized products that were created? The answer concerns what is termed *regulatory arbitrage*. The regulatory capital banks were required to keep for the tranches created from a portfolio of mortgages was less than the regulatory capital that would be required for the mortgages themselves. This is because the mortgages were kept in what is referred to as the “banking book,” whereas the tranches were kept in what is referred to as the “trading book.” Capital requirements were different for the banking book and the trading book. We will discuss this point further in later chapters.

7.4.2 *Incentives*

Economists use the term “agency costs” to describe the situation where incentives are such that the interests of two parties in a business relationship are not perfectly aligned. The process by which mortgages were originated, securitized, and sold to investors was unfortunately riddled with agency costs.

The incentive of the originators of mortgages was to make loans that would be acceptable to the creators of the ABS and ABS CDO tranches. The incentive of the individuals who valued houses on which the mortgages were written was to please the lender by providing as high a valuation as possible so that the loan-to-value ratio was as low as possible. (Pleasing the lender was likely to lead to more business from that lender.) The main concern of the creators of ABSs and ABS CDOs was the profitability of the structures (i.e., the excess of the weighted average inflows over the weighted average outflows). They wanted the volume of AAA-rated tranches that they created to be as high as possible and found ways of using the published criteria of rating agencies to achieve this. The rating agencies were paid by the issuers of the securities they rated and about half their income came from structured products.

Another source of agency costs concerns financial institutions and their employees. Employee compensation falls into three categories: regular salary, the end-of-year bonus, and stock or stock options. Many employees at all levels of seniority in financial institutions, particularly traders, receive much of their compensation in the form of end-of-year bonuses. Traditionally, this form of compensation has focused employee attention on short-term performance. If an employee generates huge profits one year and is responsible for severe losses the next year, the employee will receive a big bonus the first year and will not have to return it the following year. The employee might lose his or her job as a result of the second-year losses, but even that is not a disaster. Financial institutions seem to be surprisingly willing to recruit individuals with losses on their resumes.

Imagine you are an employee of a financial institution investing in ABS CDOs in 2006. Almost certainly you would have recognized that there was a bubble in the U.S.

housing market and would expect that bubble to burst sooner or later. However, it is possible that you would decide to continue with your ABS CDO investments. If the bubble did not burst until after December 31, 2006, you would still get a nice bonus at the end of 2006!

7.5 Lessons from the Global Financial Crisis

Some of the lessons for risk managers from the GFC are as follows:

1. Risk managers should be watching for situations where there is irrational exuberance and make sure that senior management recognize that the good times will not last forever.
2. Correlations always increase in stressed markets. In considering how bad things might get, risk managers should not use correlations that are estimated from data collected during normal market conditions.
3. Recovery rates decline when default rates increase. This is true for almost all debt instruments, not just mortgages. (See Section 17.3.) In considering how bad things might get, risk managers should not use recovery rates that are estimated from data collected during normal market conditions.
4. Risk managers should ensure that the incentives of traders and other personnel encourage them to make decisions that are in the interests of the organization they work for. Many financial institutions have revised their compensation policies as a result of the crisis. Bonuses are now often spread out over several years rather than all being paid at once. If good performance in one year is followed by bad performance in the next, part of the bonus for the good-performance year that has not yet been paid may be clawed back.
5. If a deal seems too good to be true, it probably is. AAA-rated tranches of structured products promised returns that were higher than the returns promised on AAA bonds by 100 basis points or more. A sensible conclusion from this for an investor would be that further analysis is needed because there are likely to be risks in the tranches that are not considered by rating agencies.
6. Investors should not rely on ratings. They should understand the assumptions made by rating agencies and carry out their own analyses.
7. Transparency is important in financial markets. If there is a lack of transparency (as there was for ABS CDOs), markets are liable to dry up when there is negative news.
8. Re-securitization, which led to the creation of ABS CDOs and CDOs of CDOs, was a badly flawed idea. The assets used to create ABSs in the first leg of the securitization should be as well diversified as possible. There is then nothing to be gained from further securitization.

Business Snapshot 7.1 makes the point that many market participants incorrectly considered ABS tranches rated BBB to be equivalent to BBB bonds. Business Snapshot 7.2 suggests a trading strategy that could be followed by people who realized that this was not so.

BUSINESS SNAPSHOT 7.2**A Trading Opportunity?**

A few traders made a huge amount of money betting against the subprime mortgage market. Suppose that you are analyzing markets in 2005 and 2006, but are uncertain about how subprime mortgages will perform. Is there a trading opportunity open to you?

The answer is that Mezz ABS CDOs do present a trading opportunity. Figure 7.5 is a simplification of how tranches were actually created. In practice, there were usually three ABS tranches, rated BBB+, BBB, and BBB-. Each was very thin—about 1% wide. Separate Mezz ABS CDOs were created from each of the three types of tranches. Consider the Mezz ABS CDO created from BBB+ tranches. After analyzing the mortgages in the pools used to create the BBB+ tranches, a trader would conclude that the pools had similar risk profiles and that there would be a tendency for either all tranches to be safe or all to be wiped out. (Because the tranches are only 1% wide, it is unlikely that they would be only partially wiped out.) This means that all the Mezz ABS CDO tranches created from ABS tranches rated BBB+ are either safe or wiped out. The Mezz ABS CDO tranches are therefore much the same as each other and should have the same rating (BBB+ in the case we are considering).

Having recognized this, what should the trader do? He or she should buy junior ABS CDO tranches (which are inexpensive because of their rating) and short senior ABS CDO tranches (which are relatively expensive). If the underlying principal is the same for both trades, the trader can then relax knowing that a profit has been locked in.

This emphasizes the point in Business Snapshot 7.1 that BBB tranches (particularly very thin BBB tranches) should not be considered equivalent to BBB bonds when a securitization is planned.

Summary

The GFC had a devastating effect on financial markets throughout the world. Its origins can be found in the U.S. housing market. The U.S. government was keen to encourage home ownership. Interest rates were low. Mortgage brokers and mortgage lenders found it attractive to do more business by relaxing their lending standards. Products for securitizing mortgages had been developed so that the investors bearing the credit risk were not necessarily the same as the original lenders. Rating agencies were prepared to give an AAA rating to senior tranches that were created by securitization. There was no shortage of buyers for these AAA-rated tranches because their yields were higher than the yields

on AAA-rated bonds. The compensation arrangements in banks focused their employees' attention on short-term profits, and as a result they chose to ignore the housing bubble and its potential impact on some very complex products they were trading.

House prices rose as both first-time buyers and speculators entered the market. Some mortgages had included a low "teaser rate" for two or three years. After the teaser rate ended, some borrowers faced higher interest rates that they could not afford and had no choice but to default. This led to foreclosures and an increase in the supply of houses being sold. The price increases between 2000 and 2006 began to be reversed. Speculators and others who found that the amount owed on their mortgages was greater than the value of their houses (i.e., they had negative equity) defaulted. This accentuated the price decline.

Many factors played a part in creating the U.S. housing bubble and resulting recession. These include irrational exuberance on the part of market participants, poor incentives, too much reliance on rating agencies, not enough analysis by investors, and the complexity of the products that were created. The GFC has provided a number of lessons for risk managers. As we will see later in this book, it has also led to a major overhaul of bank regulation and bank legislation.

Further Reading

- Gorton, G. "The Subprime Panic." *European Financial Management* 15, no. 1 (2008): 10–46.
- Hull, J. C. "The Financial Crisis of 2007: Another Case of Irrational Exuberance." In *The Finance Crisis and Rescue: What Went Wrong? Why? What Lessons Can Be Learned?* University of Toronto Press, 2008.
- Keys, B. J., T. Mukherjee, A. Seru, and V. Vig. "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans." *Quarterly Journal of Economics* 125, no. 1 (February 2010): 307–362.
- Krinsman, A. N. "Subprime Mortgage Meltdown: How Did It Happen and How Will It End?" *Journal of Structured Finance* (Summer 2007): 13–19.
- Mian, A., and A. Sufi. "The Consequences of Mortgage Credit Expansion: Evidence from the US Mortgage Default Crisis." *Quarterly Journal of Economics* 124, no. 4 (November 2009): 1449–1496.
- Sorkin, A. R. *Too Big to Fail*. New York: Penguin, 2009.
- Tett, G. *Fool's Gold: How the Bold Dream of a Small Tribe at JPMorgan Was Corrupted by Wall Street Greed and Unleashed a Catastrophe*. New York: Free Press, 2009.
- Zimmerman, T. "The Great Subprime Meltdown." *Journal of Structured Finance* (Fall 2007): 7–20.

Practice Questions and Problems (Answers at End of Book)

- 7.1 Why did mortgage lenders frequently not check on information in the mortgage application during the 2000 to 2007 period?
- 7.2 Why do you think the increase in house prices during the 2000 to 2007 period is referred to as a bubble?
- 7.3 What are the numbers in Table 7.1 for a loss rate of (a) 5% and (b) 12%?

- 7.4 In what ways are the risks in the tranche of an ABS different from the risks in a similarly rated bond?
- 7.5 Explain the difference between (a) an ABS and (b) an ABS CDO.
- 7.6 How were the risks in ABS CDOs misjudged by the market?
- 7.7 What is meant by the term “agency costs”?
- 7.8 What is a waterfall in a securitization?
- 7.9 How is an ABS CDO created? What was the motivation to create ABS CDOs?
- 7.10 How did Mian and Sufi show that mortgage lenders relaxed their lending criteria during the 2000 to 2006 period?
- 7.11 What is a mezzanine tranche?
- 7.12 Explain the influence of an increase in default correlation on (a) the risks in the equity tranche of an ABS and (b) the risks in the senior tranches of an ABS.
- 7.13 Explain why the end-of-year bonus has, in the past, been regarded as providing incentives for employees to think only about the short term.

Further Questions

- 7.14 Suppose that the principals assigned to the senior, mezzanine, and equity tranches are 70%, 20%, and 10% instead of 75%, 20% and 5% for both ABSs and ABS CDO in Figure 7.4. How are the results in Table 7.1 affected?
- 7.15 Investigate what happens as the width of the mezzanine tranche of the ABS in Figure 7.4 is decreased, with the reduction in the mezzanine tranche principal being divided equally between the equity and senior tranches. In particular, what is the effect on Table 7.1?

Chapter 8

Volatility

It is important for a financial institution to monitor the volatilities of the market variables (exchange rates, equity prices, commodity prices, etc.) on which the value of its portfolio depends. This chapter describes the procedures it can use to do this.

The chapter starts by explaining how volatility is defined. It then examines the common assumption that percentage returns from market variables are normally distributed and presents the power law as an alternative. After that it moves on to consider models with imposing names such as exponentially weighted moving average (EWMA), autoregressive conditional heteroscedasticity (ARCH), and generalized autoregressive conditional heteroscedasticity (GARCH). The distinctive feature of these models is that they recognize that volatility is not constant. During some periods, volatility is relatively low, while during other periods it is relatively high. The models attempt to keep track of variations in volatility through time.

8.1 Definition of Volatility

A variable's volatility, σ , is defined as the standard deviation of the return provided by the variable per unit of time when the return is expressed using continuous compounding. (See Appendix A for a discussion of compounding frequencies.) When volatility is used for option pricing, the unit of time is usually one year, so that volatility is the standard deviation of the continuously compounded return per year. When volatility is used for risk management, the unit of time is usually one day so that volatility is the standard deviation of the continuously compounded return per day.

Define S_i as the value of a variable at the end of day i . The continuously compounded return per day for the variable on day i is

$$\ln \frac{S_i}{S_{i-1}}$$

Because the time period is so short, this is almost exactly the same as

$$\frac{S_i - S_{i-1}}{S_{i-1}}$$

An alternative definition of daily volatility of a variable is therefore the standard deviation of the proportional change in the variable during a day. This is the definition that is usually used in risk management.

Example 8.1

Suppose that an asset price is \$60 and that its daily volatility is 2%. This means that a one-standard-deviation move in the asset price over one day would be 60×0.02 or \$1.20. If we assume that the change in the asset price is normally distributed, we can be 95% certain that the asset price will be between $60 - 1.96 \times 1.2 = \$57.65$ and $60 + 1.96 \times 1.2 = \$62.35$ at the end of the day. (This is what is referred to as a two-tailed test, where there is a 2.5% probability in each of the upper and lower tails of the distribution.)

If we assume that the continuously compounded returns each day are independent with the same variance, the variance of the continuously compounded return over T days is T times the variance of this return over one day. This means that the standard deviation of the return over T days is $\sqrt{T}$ times the standard deviation of the return over one day. This is consistent with the adage “uncertainty increases with the square root of time.”

Example 8.2

Assume as in Example 8.1 that an asset price is \$60 and the volatility per day is 2%. The standard deviation of the continuously compounded return over five days is $\sqrt{5} \times 2$ or 4.47%. Because five days is a short period of time, this can be assumed to be the same as the standard deviation of the proportional change over five days. A one-standard-deviation move would be $60 \times 0.0447 = 2.68$. If we assume that the change in the asset price is normally distributed, we can be 95% certain that the asset price will be between $60 - 1.96 \times 2.68 = \54.74 and $60 + 1.96 \times 2.68 = \65.26 at the end of the five days.

8.1.1 Variance Rate

Risk managers often focus on the variance rate rather than the volatility. The variance rate per day is the variance of the return in one day. As already indicated, the standard deviation of the return in time T increases with the square root of T , the variance of this return increases linearly with T . If we wanted to be pedantic, we could say that it is correct to talk about the variance rate per day, but volatility is per square root of day.

8.1.2 Business Days vs. Calendar Days

One issue is whether time should be measured in calendar days or business days. As shown in Business Snapshot 8.1, research shows that volatility is much higher on business days than on non-business days. As a result, analysts tend to ignore weekends and holidays when calculating and using volatilities. The usual assumption is that there are 252 days per year.

Assuming that the returns on successive days are independent and have the same standard deviation, this means that

$$\sigma_{\text{yr}} = \sigma_{\text{day}} \sqrt{252}$$

or

$$\sigma_{\text{day}} = \frac{\sigma_{\text{yr}}}{\sqrt{252}}$$

showing that the daily volatility is about 6% of annual volatility.

8.2 Implied Volatilities

Although risk managers usually calculate volatilities from historical data, they also try and keep track of what are known as *implied volatilities*. The one parameter in the famous Black–Scholes–Merton option pricing model that cannot be observed directly is the volatility of the underlying asset price (see Appendix E). The implied volatility of an option is the volatility that gives the market price of the option when it is substituted into this pricing model.

8.2.1 The VIX Index

The CBOE publishes indices of implied volatility. The most popular index, the VIX, is an index of the implied volatility of 30-day options on the S&P 500 calculated from a wide range of calls and puts.¹ Trading in futures on the VIX started in 2004 and trading in

¹Similarly, the VXN is an index of the volatility of the NASDAQ 100 index and the VXD is an index of the volatility of the Dow Jones Industrial Average.

BUSINESS SNAPSHOT 8.1

What Causes Volatility?

It is natural to assume that the volatility of a stock or other asset is caused by new information reaching the market. This new information causes people to revise their opinions about the value of the asset. The price of the asset changes and volatility results. However, this view of what causes volatility is not supported by research.

With several years of daily data on an asset price, researchers can calculate:

1. The variance of the asset's returns between the close of trading on one day and the close of trading on the next day when there are no intervening nontrading days.
2. The variance of the asset's return between the close of trading on Friday and the close of trading on Monday.

The second is the variance of returns over a three-day period. The first is a variance over a one-day period. We might reasonably expect the second variance to be three times as great as the first variance. Fama (1965), French (1980), and French and Roll (1986) show that this is not the case. For the assets considered, the three research studies estimate the second variance to be 22%, 19%, and 10.7% higher than the first variance, respectively.

At this stage you might be tempted to argue that these results are explained by more news reaching the market when the market is open for trading. But research by Roll (1984) does not support this explanation. Roll looked at the prices of orange juice futures. By far the most important news for orange juice futures is news about the weather, and this is equally likely to arrive at any time. When Roll compared the two variances for orange juice futures, he found that the second (Friday-to-Monday) variance is only 1.54 times the first (one-day) variance.

The only reasonable conclusion from all this is that volatility is, to a large extent, caused by trading itself. (Traders usually have no difficulty accepting this conclusion!)

options on the VIX started in 2006. A trade involving options on the S&P 500 is a bet on both the future level of the S&P 500 and the volatility of the S&P 500. By contrast, a futures or options contract on the VIX is a bet only on volatility. One contract is on 1,000 times the index.

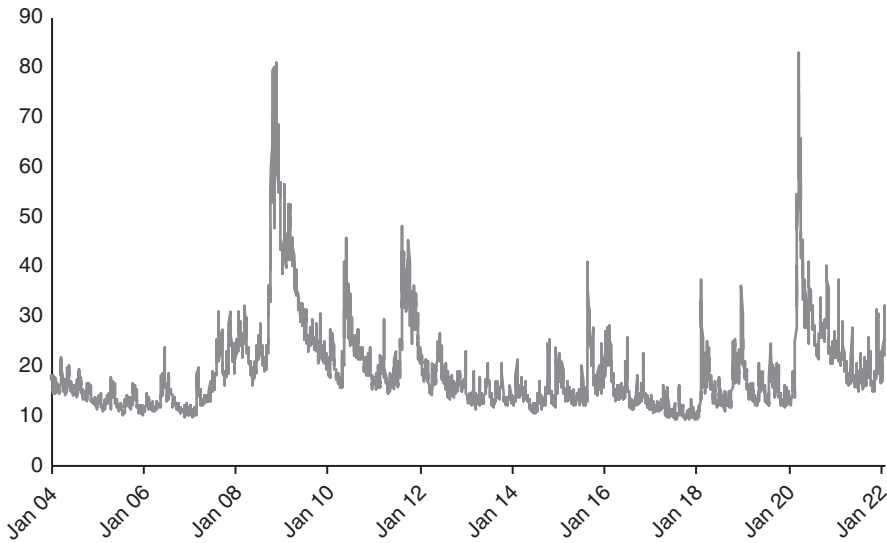


Figure 8.1 The VIX Index, January 2, 2004, to February 1, 2022

Example 8.3

Suppose that a trader buys an April futures contract on the VIX when the futures price is \$18.5 (corresponding to a 30-day S&P 500 volatility of 18.5%) and closes out the contract when the futures price is \$19.3 (corresponding to an S&P 500 volatility of 19.3%). The trader makes a gain of \$800.

Figure 8.1 shows the VIX index between January 2, 2004, and February 1, 2022. Between 2004 and mid-2007, it tended to stay between 10 and 20. It reached 30 during the second half of 2007 and 80 in October and November 2008 after the failure of Lehman Brothers. By early 2010, it had returned to more normal levels, but it spiked again during March 2020 when the beginning of the COVID-19 pandemic led to a great deal of uncertainty about the future and a big market decline. Sometimes market participants refer to the VIX index as the fear index.

8.3 Are Daily Percentage Changes in Financial Variables Normal?

When confidence limits for the change in a market variable are calculated from its volatility, a common assumption is that the change is normally distributed. This is the assumption we made in Examples 8.1 and 8.2. In practice, most financial variables are more likely

to experience big moves than the normal distribution would suggest. Table 8.1 shows the results of a test of normality using daily movements in 10 different exchange rates over a 10-year period between 2005 and 2015. The exchange rates are those between the U.S. dollar and the following currencies: Australian dollar, British pound, Canadian dollar, Danish krone, euro, Japanese yen, Mexican peso, New Zealand dollar, Swedish krona, and Swiss franc. The first step in the production of the table is to calculate the standard deviation of daily percentage changes in each exchange rate. The next stage is to note how often the actual percentage changes exceeded one standard deviation, two standard deviations, and so on. These numbers are then compared with the corresponding numbers for the normal distribution.

Daily percentage changes exceed three standard deviations on 1.30% of the days. The normal model for returns predicts that this should happen on only 0.27% of days. Daily percentage changes exceed four, five, and six standard deviations on 0.49%, 0.24%, and 0.13% of the days, respectively. The normal model predicts that we should hardly ever observe this happening. The table, therefore, provides evidence to support the existence of the fact that the probability distributions of changes in exchange rates have heavier tails than the normal distribution.

When returns are continuously compounded, the return over many days is the sum of the returns over each of the days. If the probability distribution of the return in a day were the same non-normal distribution each day, the central limit theorem of statistics would lead to the conclusion that the return over many days is normally distributed. In fact, the returns on successive days are not identically distributed. (One reason for this, which will be discussed later in this chapter, is that volatility is not constant.) As a result, heavy tails are observed in the returns over relatively long periods as well as in the returns observed over one day. Business Snapshot 8.2 shows how you could have made money if you had done an analysis similar to that in Table 8.1 in 1985!

Figure 8.2 compares a typical heavy-tailed distribution (such as the one for an exchange rate) with a normal distribution that has the same mean and standard

Table 8.1 Percentage of Days When Absolute Size of Daily Exchange Rate Moves Is Greater Than One, Two, ... , Six Standard Deviations (S.D. = standard deviation of percentage daily change)

	Real World (%)	Normal Model (%)
> 1 S.D.	23.32	31.73
> 2 S.D.	4.67	4.55
> 3 S.D.	1.30	0.27
> 4 S.D.	0.49	0.01
> 5 S.D.	0.24	0.00
> 6 S.D.	0.13	0.00

BUSINESS SNAPSHOT 8.2

Making Money from Foreign Currency Options

Black, Scholes, and Merton in their option pricing model assume that the underlying asset's price has a lognormal distribution at a future time. This is equivalent to the assumption that asset price changes over short periods, such as one day, are normally distributed. Suppose that most market participants are comfortable with the assumptions made by Black, Scholes, and Merton. You have just done the analysis in Table 8.1 and know that the normal/lognormal assumption is not a good one for exchange rates. What should you do?

The answer is that you should buy deep-out-of-the-money call and put options on a variety of different currencies—and wait. These options will be relatively inexpensive and more of them will close in-the-money than the Black–Scholes–Merton model predicts. The present value of your payoffs will on average be much greater than the cost of the options.

In the mid-1980s, a few traders knew about the heavy tails of foreign exchange probability distributions. Everyone else thought that the lognormal assumption of the Black–Scholes–Merton model was reasonable. The few traders who were well informed followed the strategy we have described—and made lots of money. By the late 1980s, most traders understood the heavy tails and the trading opportunities had disappeared.

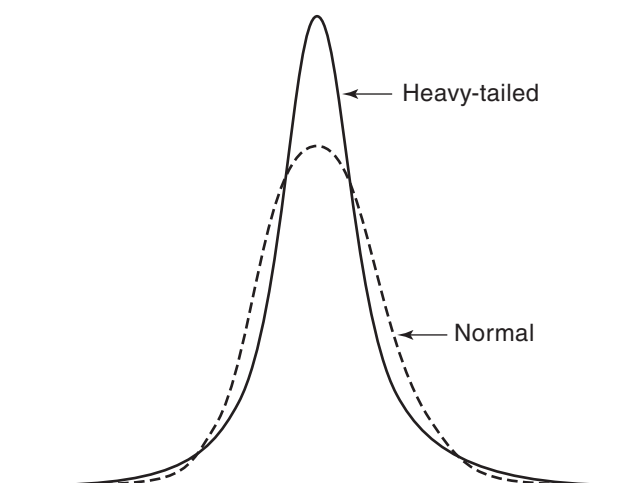


Figure 8.2 Comparison of a Normal Distribution with a Heavy-Tailed Distribution
The two distributions have the same mean and standard deviation.

deviation.² The heavy-tailed distribution is more peaked than the normal distribution. In Figure 8.2, we can distinguish three parts of the distribution: the middle, the tails, and the intermediate parts (between the middle and the tails). When we move from the normal distribution to the heavy-tailed distribution, probability mass shifts from the intermediate parts of the distribution to both the tails and the middle. If we are considering the percentage change in a market variable, the heavy-tailed distribution has the property that small and large changes in the variable are more likely than they would be if a normal distribution were assumed. Intermediate changes are less likely.

8.4 The Power Law

The power law provides an alternative to assuming normal distributions. The law asserts that, for many variables that are encountered in practice, it is approximately true that the value of the variable, ν , has the property that when x is large

$$\text{Prob}(\nu > x) = Kx^{-\alpha} \quad (8.1)$$

where K and α are constants. The equation has been found to be approximately true for variables ν as diverse as the income of an individual, the size of a city, and the number of visits to a website in a day.

Example 8.4

Suppose that we know from experience that $\alpha = 3$ for a particular financial variable and we observe that the probability that $\nu > 10$ is 0.05. Equation (8.1) gives

$$0.05 = K \times 10^{-3}$$

so that $K = 50$. The probability that $\nu > 20$ can now be calculated as

$$50 \times 20^{-3} = 0.00625$$

The probability that $\nu > 30$ is

$$50 \times 30^{-3} = 0.0019$$

and so on.

²*Kurtosis* measures the size of a distribution's tails. A *leptokurtic distribution* has heavier tails than the normal distribution. A *platykurtic distribution* has less heavy tails than the normal distribution; a distribution with the same-size tails as the normal distribution is termed *mesokurtic*.

Table 8.2 Values Calculated from Table 8.1

x	$\ln(x)$	$\text{Prob}(\nu > x)$	$\ln[\text{Prob}(\nu > x)]$
1	0.000	0.2332	-1.4560
2	0.693	0.0467	-3.0634
3	1.099	0.0130	-4.3421
4	1.386	0.0049	-5.3168
5	1.609	0.0024	-6.0182
6	1.792	0.0013	-6.6325

Equation (8.1) implies that

$$\ln[\text{Prob}(\nu > x)] = \ln K - \alpha \ln x$$

We can therefore do a quick test of whether it holds by plotting $\ln[\text{Prob}(\nu > x)]$ against $\ln(x)$. In order to do this for the data in Table 8.1, define ν as the number of standard deviations by which an exchange rate changes in one day.

The values of $\ln(x)$ and $\ln[\text{Prob}(\nu > x)]$ are calculated in Table 8.2. The data in Table 8.2 are plotted in Figure 8.3. This shows that the logarithm of the probability of the exchange rate changing by more than x standard deviations is approximately linearly dependent on $\ln(x)$ for $x \geq 2$. This is evidence that the power law holds for these data. Using data for $x = 3, 4, 5$, and 6 , a regression analysis gives the best-fit relationship as

$$\ln[\text{Prob}(\nu > x)] = -0.735 - 3.291 \ln(x)$$

showing that estimates for K and α are as follows: $K = e^{-0.735} = 0.479$ and $\alpha = 3.291$. An estimate for the probability of a change greater than 4.5 standard deviations is

$$0.479 \times 4.5^{-3.291} = 0.00340$$

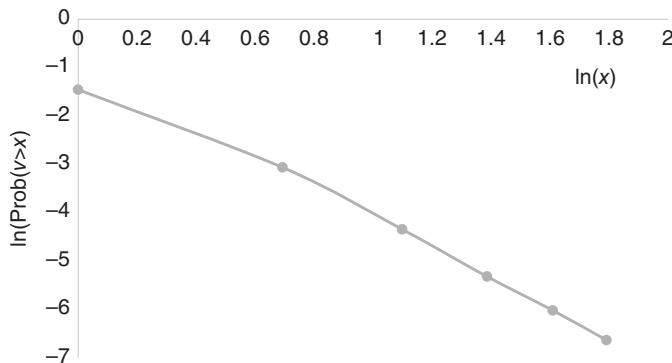


Figure 8.3 Log-Log Plot for Probability That Exchange Rate Moves by More Than a Certain Number of Standard Deviations

ν is the exchange rate change measured in standard deviations.

An estimate for the probability of a change greater than seven standard deviations is

$$0.479 \times 7^{-3.291} = 0.000794$$

We examine the power law more formally and explain better procedures for estimating the parameters when we consider extreme value theory in Chapter 12.

8.5 Monitoring Daily Volatility

Define σ_n as the volatility per day of a market variable on day n , as estimated at the end of day $n - 1$. The *variance rate*, which, as mentioned earlier, is defined as the square of the volatility, is σ_n^2 . Suppose that the value of the market variable at the end of day i is S_i . Define u_i as the continuously compounded return during day i (between the end of day $i - 1$ and the end of day i) so that

$$u_i = \ln \frac{S_i}{S_{i-1}}$$

One approach to estimating σ_n is to set it equal to the standard deviation of the u_i 's. When the most recent m observations on the u_i are used in conjunction with the usual formula for standard deviation, this approach gives:

$$\sigma_n^2 = \frac{1}{m-1} \sum_{i=1}^m (u_{n-i} - \bar{u})^2 \quad (8.2)$$

where $\bar{u}$ is the mean of the u_i 's:

$$\bar{u} = \frac{1}{m} \sum_{i=1}^m u_{n-i}$$

Example 8.5

Table 8.3 shows a possible sequence of stock prices. Suppose that we are interested in estimating the volatility for day 21 using 20 observations on the u_i so that $n = 21$ and $m = 20$. In this case, $\bar{u} = 0.00074$ and the estimate of the standard deviation of the daily return calculated using equation (8.2) is 1.49%.

For risk management purposes, the formula in equation (8.2) is usually changed in a number of ways:

1. As indicated in Section 8.1, u_i is defined as the percentage change in the market variable between the end of day $i - 1$ and the end of day i so that

$$u_i = \frac{S_i - S_{i-1}}{S_{i-1}} \quad (8.3)$$

This makes very little difference to the values of u_i that are computed.

Table 8.3 Data for Computation of Volatility

Day	Closing Stock Price (dollars)	Price Relative S_i/S_{i-1}	Daily Return $u_i = \ln(S_i/S_{i-1})$
0	20.00		
1	20.10	1.00500	0.00499
2	19.90	0.99005	-0.01000
3	20.00	1.00503	0.00501
4	20.50	1.02500	0.02469
5	20.25	0.98780	-0.01227
6	20.90	1.03210	0.03159
7	20.90	1.00000	0.00000
8	20.90	1.00000	0.00000
9	20.60	0.98565	-0.01446
10	20.50	0.99515	-0.00487
11	21.00	1.02439	0.02410
12	21.10	1.00476	0.00475
13	20.70	0.98104	-0.01914
14	20.50	0.99034	-0.00971
15	20.70	1.00976	0.00971
16	20.90	1.00966	0.00962
17	20.40	0.97608	-0.02421
18	20.50	1.00490	0.00489
19	20.60	1.00488	0.00487
20	20.30	0.98544	-0.01467

2. $\bar{u}$ is assumed to be zero. The justification for this is that the expected change in a variable in one day is very small when compared with the standard deviation of changes.³
3. $m - 1$ is replaced by m . This moves us from an unbiased estimate of the volatility to a maximum likelihood estimate, as we explain in Section 8.9.

These three changes allow the formula for the variance rate to be simplified to

$$\sigma_n^2 = \frac{1}{m} \sum_{i=1}^m u_{n-i}^2 \quad (8.4)$$

where u_i is given by equation (8.3).

³This is likely to be the case even if the variable happened to increase or decrease quite fast during the m days of our data.

Example 8.6

Consider again Example 8.5. When $n = 21$ and $m = 20$,

$$\sum_{i=1}^m u_{n-i}^2 = 0.00424$$

so that equation (8.4) gives

$$\sigma_n^2 = 0.00424/20 = 0.000214$$

and $\sigma_n = 0.014618$ or 1.46%. This is only a little different from the result in Example 8.5.

8.5.1 Weighting Schemes

Equation (8.4) gives equal weight to each of u_{n-1}^2 , u_{n-2}^2 , ..., and u_{n-m}^2 . Our objective is to estimate σ_n , the volatility on day n . It therefore makes sense to give more weight to recent data. A model that does this is

$$\sigma_n^2 = \sum_{i=1}^m \alpha_i u_{n-i}^2 \quad (8.5)$$

The variable α_i is the amount of weight given to the observation i days ago. The α 's are positive. If we choose them so that $\alpha_i < \alpha_j$ when $i > j$, less weight is given to older observations. The weights must sum to unity, so that

$$\sum_{i=1}^m \alpha_i = 1$$

An extension of the idea in equation (8.5) is to assume that there is a long-run average variance rate and that this should be given some weight. This leads to the model that takes the form

$$\sigma_n^2 = \gamma V_L + \sum_{i=1}^m \alpha_i u_{n-i}^2 \quad (8.6)$$

where V_L is the long-run variance rate and γ is the weight assigned to V_L . Because the weights must sum to unity, we have

$$\gamma + \sum_{i=1}^m \alpha_i = 1$$

This is known as an ARCH(m) model. It was first suggested by Engle.⁴ The estimate of the variance is based on a long-run average variance and m observations. The older an observation, the less weight it is given. Defining $\omega = \gamma V_L$, the model in equation (8.6) can be written

$$\sigma_n^2 = \omega + \sum_{i=1}^m \alpha_i u_{n-i}^2 \quad (8.7)$$

In the next two sections, we discuss two important approaches to monitoring volatility using the ideas in equations (8.5) and (8.6).

8.6 The Exponentially Weighted Moving Average Model

The exponentially weighted moving average (EWMA) model is a particular case of the model in equation (8.5) where the weights, α_i , decrease exponentially as we move back through time. Specifically, $\alpha_{i+1} = \lambda \alpha_i$ where λ is a constant between zero and one.

It turns out that this weighting scheme leads to a particularly simple formula for updating volatility estimates. The formula is

$$\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1 - \lambda) u_{n-1}^2 \quad (8.8)$$

The estimate, σ_n , of the volatility for day n (made at the end of day $n - 1$) is calculated from σ_{n-1} (the estimate that was made at the end of day $n - 2$ of the volatility for day $n - 1$) and u_{n-1} (the most recent daily percentage change).

To understand why equation (8.8) corresponds to weights that decrease exponentially, we substitute for σ_{n-1}^2 to get

$$\sigma_n^2 = \lambda[\lambda \sigma_{n-2}^2 + (1 - \lambda) u_{n-2}^2] + (1 - \lambda) u_{n-1}^2$$

or

$$\sigma_n^2 = (1 - \lambda)(u_{n-1}^2 + \lambda u_{n-2}^2) + \lambda^2 \sigma_{n-2}^2$$

Substituting in a similar way for σ_{n-2}^2 gives

$$\sigma_n^2 = (1 - \lambda)(u_{n-1}^2 + \lambda u_{n-2}^2 + \lambda^2 u_{n-3}^2) + \lambda^3 \sigma_{n-3}^2$$

⁴See R. F. Engle, "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica* 50 (1982): 987–1008. Robert Engle won the Nobel Prize in Economics in 2003 for his work on ARCH models.

Continuing in this way, we see that

$$\sigma_n^2 = (1 - \lambda) \sum_{i=1}^m \lambda^{i-1} u_{n-i}^2 + \lambda^m \sigma_{n-m}^2$$

For a large m , the term $\lambda^m \sigma_{n-m}^2$ is sufficiently small to be ignored so that equation (8.8) is the same as equation (8.5) with $\alpha_i = (1 - \lambda)\lambda^{i-1}$. The weights for the u_i decline at rate λ as we move back through time. Each weight is λ times the previous weight.

Example 8.7

Suppose that λ is 0.90, the volatility estimated for a market variable for day $n - 1$ is 1% per day, and during day $n - 1$ the market variable increased by 2%. This means that $\sigma_{n-1}^2 = 0.01^2 = 0.0001$ and $u_{n-1}^2 = 0.02^2 = 0.0004$. Equation (8.8) gives

$$\sigma_n^2 = 0.9 \times 0.0001 + 0.1 \times 0.0004 = 0.00013$$

The estimate of the volatility for day n , σ_n , is, therefore, $\sqrt{0.00013}$ or 1.14% per day. Note that the expected value of u_{n-1}^2 is σ_{n-1}^2 or 0.0001. In this example, the realized value of u_{n-1}^2 is greater than the expected value, and as a result our volatility estimate increases. If the realized value of u_{n-1}^2 had been less than its expected value, our estimate of the volatility would have decreased.

The EWMA approach has the attractive feature that the data storage requirements are modest. At any given time, we need to remember only the current estimate of the variance rate and the most recent observation on the value of the market variable. When we get a new observation on the value of the market variable, we calculate a new daily percentage change and use equation (8.8) to update our estimate of the variance rate. The old estimate of the variance rate and the old value of the market variable can then be discarded.

The EWMA approach is designed to track changes in the volatility. Suppose there is a big move in the market variable on day $n - 1$ so that u_{n-1}^2 is large. From equation (8.8) this causes our estimate of the current volatility to move upward. The value of λ governs how responsive the estimate of the daily volatility is to the most recent daily percentage change. A low value of λ leads to a great deal of weight being given to the u_{n-1}^2 when σ_n is calculated. In this case, the estimates produced for the volatility on successive days are themselves highly volatile. A high value of λ (i.e., a value close to 1.0) produces estimates of the daily volatility that respond relatively slowly to new information provided by the daily percentage change.

The RiskMetrics database, which was originally created by JPMorgan and made publicly available in 1994, used the EWMA model with $\lambda = 0.94$ for updating daily volatility estimates. The company found that, across a range of different market variables,

this value of λ gives forecasts of the variance rate that come closest to the realized variance rate.⁵ RiskMetrics, now a separate company from JPMorgan, has since switched to more sophisticated models.

8.7 The GARCH(1,1) Model

We now move on to discuss what is known as the GARCH(1,1) model, proposed by Bollerslev in 1986.⁶ The difference between the EWMA model and the GARCH(1,1) model is analogous to the difference between equation (8.5) and equation (8.6). In GARCH(1,1), σ_n^2 is calculated from a long-run average variance rate, V_L , as well as from σ_{n-1} and u_{n-1} . The equation for GARCH(1,1) is

$$\sigma_n^2 = \gamma V_L + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2 \quad (8.9)$$

where γ is the weight assigned to V_L , α is the weight assigned to u_{n-1}^2 , and β is the weight assigned to σ_{n-1}^2 . Because the weights must sum to one:

$$\gamma + \alpha + \beta = 1$$

The EWMA model is a particular case of GARCH(1,1) where $\gamma = 0$, $\alpha = 1 - \lambda$, and $\beta = \lambda$.

The “(1,1)” in GARCH(1,1) indicates that σ_n^2 is based on the most recent observation of u^2 and the most recent estimate of the variance rate. The more general GARCH(p, q) model calculates σ_n^2 from the most recent p observations on u^2 and the most recent q estimates of the variance rate.⁷ GARCH(1,1) is by far the most popular of the GARCH models.

Setting $\omega = \gamma V_L$, the GARCH(1,1) model can also be written

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2 \quad (8.10)$$

⁵See JPMorgan, *RiskMetrics Monitor*, Fourth Quarter 1995. We will explain an alternative (maximum likelihood) approach to estimating parameters later in the chapter. The realized variance rate on a particular day was calculated as an equally weighted average of the u_i^2 on the subsequent 25 days. (See Problem 8.20.)

⁶See T. Bollerslev, “Generalized Autoregressive Conditional Heteroscedasticity,” *Journal of Econometrics* 31 (1986): 307–327.

⁷Other GARCH models have been proposed that incorporate asymmetric news. These models are designed so that σ_n depends on the sign of u_{n-1} . Arguably, the models are more appropriate than GARCH(1,1) for equities. This is because the volatility of an equity’s price tends to be inversely related to the price so that a negative u_{n-1} should have a bigger effect on σ_n than the same positive u_{n-1} . For a discussion of models for handling asymmetric news, see D. Nelson, “Conditional Heteroscedasticity and Asset Returns: A New Approach,” *Econometrica* 59 (1990): 347–370, and R. F. Engle and V. Ng, “Measuring and Testing the Impact of News on Volatility,” *Journal of Finance* 48 (1993): 1749–1778.

This is the form of the model that is usually used for the purposes of estimating the parameters. Once ω , α , and β have been estimated, we can calculate γ as $1 - \alpha - \beta$. The long-term variance V_L can then be calculated as ω/γ . For a stable GARCH(1,1) model, we require $\alpha + \beta < 1$. Otherwise the weight applied to the long-term variance is negative.

Example 8.8

Suppose that a GARCH(1,1) model is estimated from daily data as

$$\sigma_n^2 = 0.000002 + 0.13u_{n-1}^2 + 0.86\sigma_{n-1}^2$$

This corresponds to $\alpha = 0.13$, $\beta = 0.86$, and $\omega = 0.000002$. Because $\gamma = 1 - \alpha - \beta$, it follows that $\gamma = 0.01$ and because $\omega = \gamma V_L$, it follows that $V_L = 0.0002$. In other words, the long-run average variance per day implied by the model is 0.0002. This corresponds to a volatility of $\sqrt{0.0002} = 0.014$ or 1.4% per day.

Suppose that the estimate of the volatility on day $n - 1$ is 1.6% per day so that $\sigma_{n-1}^2 = 0.016^2 = 0.000256$, and that on day $n - 1$ the market variable decreased by 1% so that $u_{n-1}^2 = 0.01^2 = 0.0001$. Then:

$$\sigma_n^2 = 0.000002 + 0.13 \times 0.0001 + 0.86 \times 0.000256 = 0.00023516$$

The new estimate of the volatility is, therefore, $\sqrt{0.00023516} = 0.0153$ or 1.53% per day.

8.7.1 The Weights

Substituting for σ_{n-1}^2 in equation (8.10) we obtain

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta(\omega + \alpha u_{n-2}^2 + \beta \sigma_{n-2}^2)$$

or

$$\sigma_n^2 = \omega + \beta\omega + \alpha u_{n-1}^2 + \alpha\beta u_{n-2}^2 + \beta^2 \sigma_{n-2}^2$$

Substituting for σ_{n-2}^2 we get

$$\sigma_n^2 = \omega + \beta\omega + \beta^2\omega + \alpha u_{n-1}^2 + \alpha\beta u_{n-2}^2 + \alpha\beta^2 u_{n-3}^2 + \beta^3 \sigma_{n-3}^2$$

Continuing in this way, we see that the weight applied to u_{n-i}^2 is $\alpha\beta^{i-1}$. The weights decline exponentially at rate β . The parameter β can be interpreted as a decay rate. It is similar to λ in the EWMA model. It defines the relative importance of the observations on the u_i in determining the current variance rate. For example, if $\beta = 0.9$, u_{n-2}^2 is only

90% as important as u_{n-1}^2 ; u_{n-3}^2 is 81% as important as u_{n-1}^2 ; and so on. The GARCH(1,1) model is the same as the EWMA model except that, in addition to assigning weights that decline exponentially to past u_i^2 , it also assigns some weight to the long-run average variance rate.

8.8 Choosing between the Models

In practice, variance rates do tend to be pulled back to a long-run average level. The GARCH(1,1) model incorporates mean reversion whereas the EWMA model does not. GARCH(1,1), although more complicated than EWMA, is therefore theoretically more appealing than EWMA.

In the next section, we will discuss how best-fit parameters ω , α , and β in GARCH(1,1) can be estimated. When the parameter ω is zero, the GARCH(1,1) reduces to EWMA. In circumstances where the best-fit value of ω turns out to be negative, the GARCH(1,1) model is not stable and it makes sense to switch to the EWMA model.

8.9 Maximum Likelihood Methods

It is now appropriate to discuss how the parameters in the models we have been considering are estimated from historical data. The approach used is known as the *maximum likelihood method*. It involves choosing values for the parameters that maximize the chance (or likelihood) of the data occurring.

We start with a very simple example. Suppose that we sample 10 stocks at random on a certain day and find that the price of one of them declined during the day and the prices of the other nine either remained the same or increased. What is our best estimate of the proportion of stock prices that declined during the day? The natural answer is 0.1. Let us see if this is the result given by maximum likelihood methods.

Suppose that the probability of a price decline is p . The probability that one particular stock declines in price and the other nine do not is $p(1 - p)^9$. (There is a probability p that it will decline and $1 - p$ that each of the other nine will not.) Using the maximum likelihood approach, the best estimate of p is the one that maximizes $p(1 - p)^9$. Differentiating this expression with respect to p and setting the result equal to zero, it can be shown that $p = 0.1$ maximizes the expression. The maximum likelihood estimate of p is therefore 0.1, as expected.

8.9.1 Estimating a Constant Variance

In our next example of maximum likelihood methods, we consider the problem of estimating a variance of a variable X from m observations on X when the underlying distribution is normal. We assume that the observations are $u_1, u_2, \dots, u_m$ and that

the mean of the underlying normal distribution is zero. Denote the variance by ν . The likelihood of u_i being observed is the probability density function for X when $X = u_i$. This is

$$\frac{1}{\sqrt{2\pi\nu}} \exp\left(\frac{-u_i^2}{2\nu}\right)$$

The likelihood of m observations occurring in the order in which they are observed is

$$\prod_{i=1}^m \left[\frac{1}{\sqrt{2\pi\nu}} \exp\left(\frac{-u_i^2}{2\nu}\right) \right] \quad (8.11)$$

Using the maximum likelihood method, the best estimate of ν is the value that maximizes this expression.

Maximizing an expression is equivalent to maximizing the logarithm of the expression. Taking logarithms of the expression in equation (8.11) and ignoring constant multiplicative factors, it can be seen that we wish to maximize

$$\sum_{i=1}^m \left[-\ln(\nu) - \frac{u_i^2}{\nu} \right] \quad (8.12)$$

or

$$-m \ln(\nu) - \sum_{i=1}^m \frac{u_i^2}{\nu}$$

Differentiating this expression with respect to ν and setting the result equation to zero, it can be shown that the maximum likelihood estimator of ν is

$$\frac{1}{m} \sum_{i=1}^m u_i^2$$

This maximum likelihood estimator is the one we used in equation (8.4). The corresponding unbiased estimator is the same with m replaced by $m - 1$.

8.9.2 Estimating EWMA or GARCH(1,1)

We now consider how the maximum likelihood method can be used to estimate the parameters when EWMA, GARCH(1,1), or some other volatility updating procedure is used. Define $\nu_i = \sigma_i^2$ as the variance estimated for day i . Assume that the

Table 8.4 Estimation of Parameters in GARCH(1,1) Model for S&P 500 between February 2, 2017, and February 1, 2022

Date	Day i	S_i	u_i	$v_i = \sigma_i^2$	$-\ln(v_i) - u_i^2/v_i$
Feb. 02, 2017	1	2,280.85			
Feb. 03, 2017	2	2,297.42	0.007265		
Feb. 06, 2017	3	2,292.56	-0.00212	5.28E-05	9.765
Feb. 07, 2017	4	2,293.08	0.000227	4.51E-05	10.006
Feb. 08, 2017	5	2,294.67	0.000693	3.83E-05	10.158
Feb. 09, 2017	6	2,307.87	0.005752	3.32E-05	9.316
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
Jan. 31, 2022	1258	4,515.55	0.018886	2.02E-04	6.740
Feb. 01, 2022	1259	4,546.54	0.006863	2.33E-04	8.163
GARCH parameters					
ω	α	β			
0.000003914	0.2111	0.7623			

probability distribution of u_i conditional on the variance is normal. A similar analysis to the one just given shows the best parameters are the ones that maximize

$$\prod_{i=1}^m \left[\frac{1}{\sqrt{2\pi v_i}} \exp \left(\frac{-u_i^2}{2v_i} \right) \right]$$

Taking logarithms, we see that this is equivalent to maximizing

$$\sum_{i=1}^m \left[-\ln(v_i) - \frac{u_i^2}{v_i} \right] \quad (8.13)$$

This is the same as the expression in equation (8.12), except that v is replaced by v_i . It is necessary to search iteratively to find the parameters in the model that maximize the expression in equation (8.13).

The spreadsheet in Table 8.4 indicates how the calculations could be organized for the GARCH(1,1) model. The table analyzes data on the S&P 500 between February 2, 2017, and February 1, 2022.⁸

The numbers in the table are based on the estimates of the three GARCH(1,1) parameters: ω , α , and β . The first column in the table records the date. The second column counts the days. The third column shows the S&P 500 at the end of day i , S_i . The fourth column shows the proportional change in the S&P 500 between the end of day $i - 1$ and the end of day i . This is $u_i = (S_i - S_{i-1})/S_{i-1}$. The fifth column shows the estimate of the variance rate, $v_i = \sigma_i^2$, for day i made at the end of day $i - 1$. On day

⁸The data and calculations can be found at www-2.rotman.utoronto.ca/~hull/riskman.

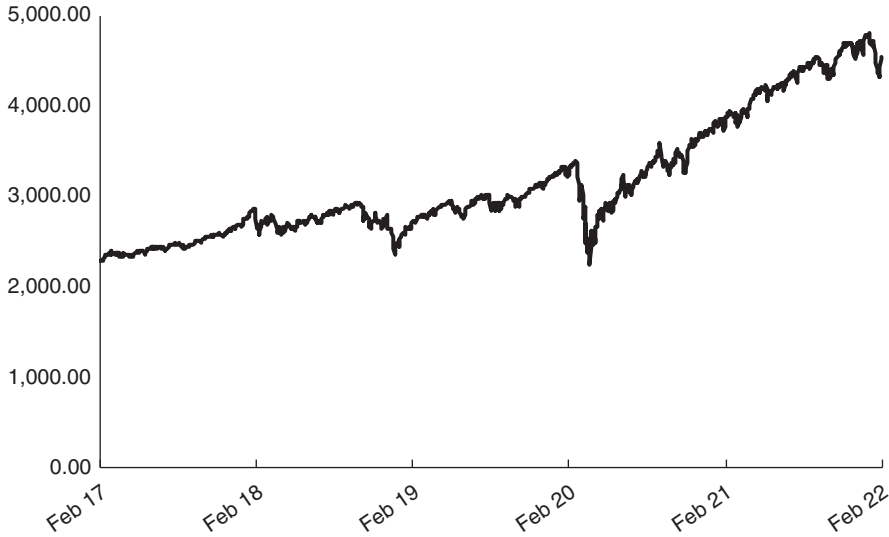


Figure 8.4 S&P 500 Index: February 2, 2017, to February 1, 2022

three, we start things off by setting the variance equal to u_2^2 . On subsequent days, equation (8.10) is used. The sixth column tabulates the likelihood measure, $-\ln(v_i) - u_i^2/v_i$. The values in the fifth and sixth columns are based on the current trial estimates of ω , α , and β . We are interested in choosing ω , α , and β to maximize the sum of the numbers in the sixth column. This involves an iterative search procedure.⁹

In our example, the optimal values of the parameters turn out to be

$$\omega = 0.000003914$$

$$\alpha = 0.2111$$

$$\beta = 0.7623$$

and the maximum value of the function in equation (8.13) is 10764.3624. (The numbers shown in Table 8.4 are those calculated on the final iteration of the search for the optimal ω , α , and β .)

The long-term variance rate, V_L , in our example is

$$\frac{\omega}{1 - \alpha - \beta} = \frac{0.000003914}{0.02662} = 0.000147$$

The long-term volatility is $\sqrt{0.000147}$ or 1.213% per day.

Figures 8.4 and 8.5 show the S&P 500 index and the GARCH(1,1) volatility during the five-year period covered by the data. Most of the time, the volatility was less

⁹As discussed later, a general-purpose algorithm such as Solver in Microsoft's Excel can be used.

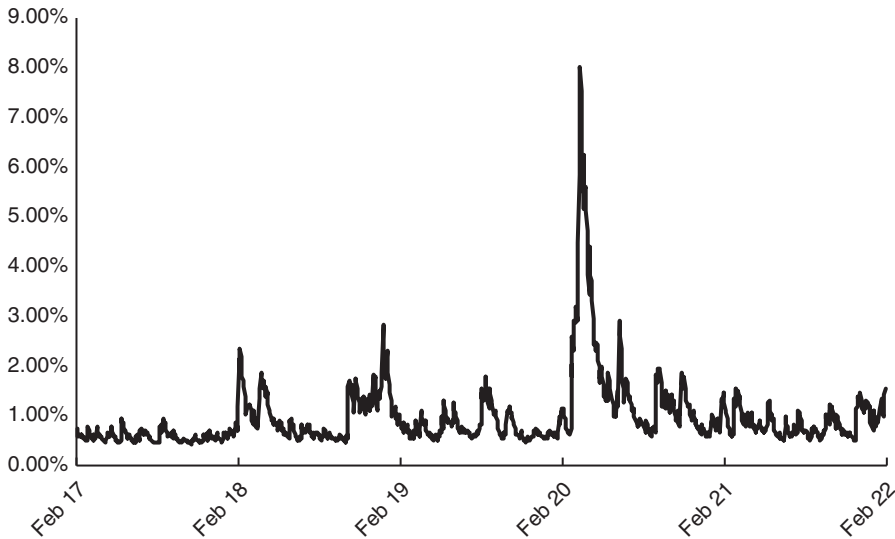


Figure 8.5 GARCH(1,1) Daily Volatility of S&P 500 Index: February 2, 2017, to February 1, 2022

than 2% per day, but occasionally the volatility spiked. The most notable spike was in March 2020 and reflected uncertainties associated with the COVID-19 pandemic. (As Figure 8.1 shows, the VIX index also spiked at that time.)

An alternative, more robust approach to estimating parameters in GARCH(1,1) is known as *variance targeting*.¹⁰ This involves setting the long-run average variance rate, V_L , equal to the sample variance calculated from the data (or to some other value that is believed to be reasonable). The value of ω then equals $V_L(1 - \alpha - \beta)$ and only two parameters have to be estimated. For the data in Table 8.4, the sample volatility is 1.2220, only a little different from the GARCH long-term volatility of 1.213%. Setting V_L equal to the sample variance (0.000149), the values of α and β that maximize the objective function in equation (8.13) are 0.2115 and 0.7622, respectively. These are very close to those using the full GARCH procedure. This indicates that variance targeting works very well in this case.

When the EWMA model is used, the estimation procedure is relatively simple. We set $\omega = 0$, $\alpha = 1 - \lambda$, and $\beta = \lambda$, and only one parameter, λ , has to be estimated. In the data in Table 8.4, the value of λ that maximizes the objective function in equation (8.13) is 0.9086 and the value of the objective function is 10,650, less than the 10,764 obtained with GARCH (1,1).

For both GARCH(1,1) and EWMA, we can use the Solver routine in Excel to search for the values of the parameters that maximize the likelihood function. The routine works well provided we structure our spreadsheet so that the parameters we are searching for have roughly equal values. For example, in GARCH(1,1) we could let cells A1, A2, and A3 contain $\omega \times 10^5$, α , and β . We could then set B1=A1/100,000. We would then use

¹⁰See R. Engle and J. Mezrich, “GARCH for Groups,” *Risk* (August 1996): 36–40.

B1, A2, and A3 for the calculations, but we would ask Solver to calculate the values of A1, A2, and A3 that maximize the likelihood function. Occasionally, Solver gives a local maximum, so a number of different starting values for the parameter should be tested.

8.9.3 How Good Is the Model?

The assumption underlying a GARCH model is that volatility changes with the passage of time. During some periods, volatility is relatively high; during other periods, it is relatively low. To put this another way, when u_i^2 is high, there is a tendency for $u_{i+1}^2, u_{i+2}^2, \dots$ to be high; when u_i^2 is low, there is a tendency for $u_{i+1}^2, u_{i+2}^2, \dots$ to be low. We can test how true this is by examining the autocorrelation structure of the u_i^2 .

Let us assume that the u_i^2 do exhibit autocorrelation. If a GARCH model is working well, it should remove the autocorrelation. We can test whether it has done this by considering the autocorrelation structure for the variables u_i^2/σ_i^2 . If these show very little autocorrelation, our model for σ_i has succeeded in explaining autocorrelations in the u_i^2 .

Table 8.5 shows results for the S&P 500 data. The first column shows the lags considered when the autocorrelation is calculated. The second column shows autocorrelations for u_i^2 ; the third column shows autocorrelations for u_i^2/σ_i^2 .¹¹ The table shows that the autocorrelations are positive for u_i^2 for all lags between 1 and 15. In the case of u_i^2/σ_i^2 , some of the autocorrelations are positive and some are negative. They are much smaller in magnitude than the autocorrelations for u_i^2 .

Table 8.5 Autocorrelations Before and After the Use of a GARCH Model

Time Lag	Autocorr for u_i^2	Autocorr for u_i^2/σ_i^2
1	0.535	0.005
2	0.557	0.006
3	0.351	0.004
4	0.349	0.040
5	0.334	-0.022
6	0.415	0.013
7	0.326	-0.016
8	0.353	-0.038
9	0.294	-0.020
10	0.259	0.057
11	0.232	-0.021
12	0.169	-0.026
13	0.171	0.012
14	0.168	-0.002
15	0.202	0.011

¹¹For a series x_i , the autocorrelation with a lag of k is the coefficient of correlation between x_i and x_{i+k} .

The GARCH model appears to have done a good job in explaining the data. For a more scientific test, we can use what is known as the Ljung-Box statistic.¹² If a certain series has m observations the Ljung-Box statistic is

$$m \sum_{k=1}^K w_k c_k^2$$

where c_k is the autocorrelation for a lag of k , K is the number of lags considered, and

$$w_k = \frac{m+2}{m-k}$$

For $K = 15$, zero autocorrelation can be rejected with 95% confidence when the Ljung-Box statistic is greater than 25.

From Table 8.5, the Ljung-Box statistic for the u_i^2 series is about 2,141. This is strong evidence of autocorrelation. For the u_i^2/σ_i^2 series, the Ljung-Box statistic is 11.5, suggesting that the autocorrelation has been largely removed by the GARCH model.

8.10 Using GARCH(1,1) to Forecast Future Volatility

The variance rate estimated at the end of day $n - 1$ for day n , when GARCH(1,1) is used, is

$$\sigma_n^2 = (1 - \alpha - \beta)V_L + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2$$

so that

$$\sigma_n^2 - V_L = \alpha(u_{n-1}^2 - V_L) + \beta(\sigma_{n-1}^2 - V_L)$$

On day $n + t$ in the future, we have

$$\sigma_{n+t}^2 - V_L = \alpha(u_{n+t-1}^2 - V_L) + \beta(\sigma_{n+t-1}^2 - V_L)$$

The expected value of u_{n+t-1}^2 is that of σ_{n+t-1}^2 . Hence,

$$E[\sigma_{n+t}^2 - V_L] = (\alpha + \beta)E[\sigma_{n+t-1}^2 - V_L]$$

¹²See G. M. Ljung and G. E. P. Box, "On a Measure of Lack of Fit in Time Series Models," *Biometrika* 65 (1978): 297–303.

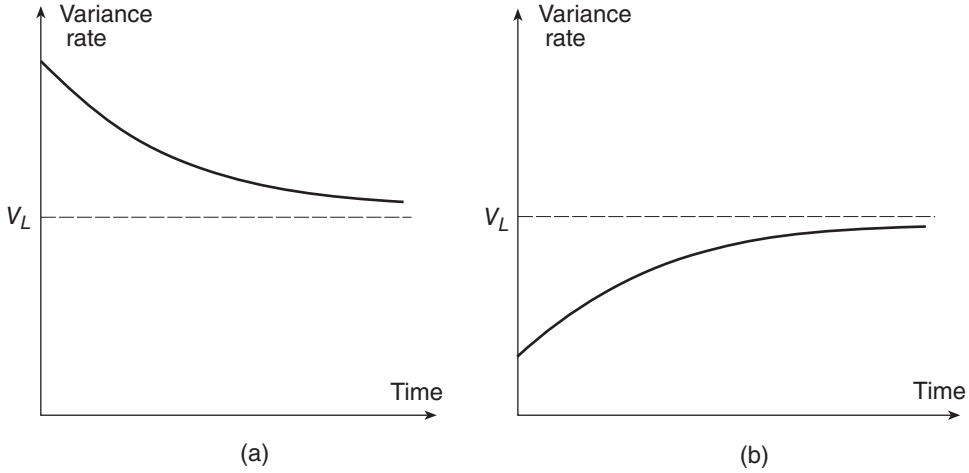


Figure 8.6 Expected Path for the Variance Rate When (a) Current Variance Rate Is Above Long-Term Variance Rate and (b) Current Variance Rate Is Below Long-Term Variance Rate

where E denotes expected value. Using this equation repeatedly yields

$$E[\sigma_{n+t}^2 - V_L] = (\alpha + \beta)^t (\sigma_n^2 - V_L)$$

or

$$E[\sigma_{n+t}^2] = V_L + (\alpha + \beta)^t (\sigma_n^2 - V_L) \quad (8.14)$$

This equation forecasts the volatility on day $n + t$ using the information available at the end of day $n - 1$. In the EWMA model, $\alpha + \beta = 1$ and equation (8.14) show that the expected future variance rate equals the current variance rate. When $\alpha + \beta < 1$, the final term in the equation becomes progressively smaller as t increases. Figure 8.6 shows the expected path followed by the variance rate for situations where the current variance rate is different from V_L . As mentioned earlier, the variance rate exhibits mean reversion with a reversion level of V_L and a reversion rate of $1 - \alpha - \beta$. Our forecast of the future variance rate tends toward V_L as we look further and further ahead. This analysis emphasizes the point that we must have $\alpha + \beta < 1$ for a stable GARCH(1,1) process. When $\alpha + \beta > 1$, the weight given to the long-term average variance is negative and the process is mean fleeing rather than mean reverting.

For the S&P 500 data considered earlier, $\alpha + \beta = 0.9734$ and $V_L = 0.0001470$. Suppose that our estimate of the current variance rate per day is 0.0003. (This corresponds to a volatility of 1.732% per day.) In 10 days, the expected variance rate is

$$0.0001470 + 0.9734^{10}(0.0003 - 0.0001470) = 0.0002638$$

The expected volatility per day is $\sqrt{0.0002638}$ or 1.62%, still well above the long-term volatility of 1.21% per day. However, the expected variance rate in 100 days is

$$0.0001470 + 0.9734^{100}(0.0003 - 0.0001470) = 0.0001573$$

and the expected volatility per day is 1.25%, quite close to the long-term volatility of 1.21%.

8.10.1 Volatility Term Structures

Suppose it is day n . Define

$$V(t) = E(\sigma_{n+t}^2)$$

and

$$a = \ln \frac{1}{\alpha + \beta}$$

so that equation (8.14) becomes

$$V(t) = V_L + e^{-at}[V(0) - V_L]$$

$V(t)$ is an estimate of the instantaneous variance rate in t days. The average variance rate per day between today and time T is

$$\frac{1}{T} \int_0^T V(t) dt = V_L + \frac{1 - e^{-aT}}{aT} [V(0) - V_L]$$

As T increases, this approaches V_L . Define $\sigma(T)$ as the volatility per annum that should be used to price a T -day option under GARCH(1,1). Assuming 252 days per year, $\sigma(T)^2$ is 252 times the average variance rate per day, so that

$$\sigma(T)^2 = 252 \left\{ V_L + \frac{1 - e^{-aT}}{aT} [V(0) - V_L] \right\} \quad (8.15)$$

The relationship between the volatilities of options and their maturities is referred to as the *volatility term structure*. The volatility term structure is usually calculated from implied volatilities, but equation (8.15) provides an alternative approach for estimating it from the GARCH(1,1) model. Although the volatility term structure estimated from GARCH(1,1) is not the same as that calculated from implied volatilities, it is often used to predict the way that the actual volatility term structure will respond to volatility changes.

Table 8.6 S&P 500 Volatility Term Structure Predicted from GARCH(1,1)

Option life (days)	10	30	50	100	500
Option volatility (% per annum)	26.62	25.20	24.13	22.45	19.98

When the current volatility is above the long-term volatility, the GARCH(1,1) model estimates a downward-sloping volatility term structure. When the current volatility is below the long-term volatility, it estimates an upward-sloping volatility term structure. In the case of the S&P 500 data, $a = \ln(1/0.97338) = 0.02698$ and $V_L = 0.0001470$. Suppose that the current variance rate per day, $V(0)$, is estimated as 0.0003 per day. It follows from equation (8.15) that

$$\sigma(T)^2 = 252 \left[0.0001470 + \frac{1 - e^{-0.02698T}}{0.02698T} (0.0003 - 0.0001470) \right]$$

where T is measured in days. Table 8.6 shows the volatility per year for different values of T .

8.10.2 Impact of Volatility Changes

Equation (8.15) can be written as

$$\sigma(T)^2 = 252 \left\{ V_L + \frac{1 - e^{-aT}}{aT} \left(\frac{\sigma(0)^2}{252} - V_L \right) \right\}$$

When $\sigma(0)$ changes by $\Delta\sigma(0)$, $\sigma(T)$ changes by approximately

$$\frac{1 - e^{-aT}}{aT} \frac{\sigma(0)}{\sigma(T)} \Delta\sigma(0) \quad (8.16)$$

Table 8.7 shows the effect of a volatility change on options of varying maturities for our S&P 500 data. We assume as before that $V(0) = 0.0003$ so that $\sigma(0) = \sqrt{252 \times 0.0003} = 27.50\%$. The table considers a 100-basis-point change in the instantaneous volatility from 27.50% per year to 28.50% per year. This means that $\Delta\sigma(0) = 0.01$ or 1%.

Many financial institutions use analyses such as this when determining the exposure of their books to volatility changes. Rather than consider an across-the-board increase of 1% in implied volatilities when calculating vega, they relate the size of the volatility increase to the maturity of the option. Based on Table 8.7, a 0.91% volatility increase would be considered for a 10-day option, a 0.75% increase for a 30-day option, a 0.63% increase for a 50-day option, and so on.

Table 8.7 Impact of 1% Increase in the Instantaneous Volatility Predicted from GARCH(1,1)

Option life (days)	10	30	50	100	500
Increase in volatility (%)	0.91	0.75	0.63	0.42	0.10

Summary

In risk management, the daily volatility of a market variable is defined as the standard deviation of the percentage daily change in the market variable. The daily variance rate is the square of the daily volatility. Volatility tends to be much higher on trading days than on nontrading days. As a result, nontrading days are ignored in volatility calculations. It is tempting to assume that daily changes in market variables are normally distributed. In fact, this is far from true. Most market variables have distributions for percentage daily changes with much heavier tails than the normal distribution. The power law has been found to be a good description of the tails of many distributions that are encountered in practice.

This chapter has discussed methods for attempting to keep track of the current level of volatility. Define u_i as the percentage change in a market variable between the end of day $i - 1$ and the end of day i . The variance rate of the market variable (that is, the square of its volatility) is calculated as a weighted average of the u_i^2 . The key feature of the methods that have been discussed here is that they do not give equal weight to the observations on the u_i^2 . The more recent an observation, the greater the weight assigned to it. In the EWMA model and the GARCH(1,1) model, the weights assigned to observations decrease exponentially as the observations become older. The GARCH(1,1) model differs from the EWMA model in that some weight is also assigned to the long-run average variance rate. Both the EWMA and GARCH(1,1) models have structures that enable forecasts of the future level of variance rate to be produced relatively easily.

Maximum likelihood methods are usually used to estimate parameters in GARCH(1,1) and similar models from historical data. These methods involve using an iterative procedure to determine the parameter values that maximize the chance or likelihood that the historical data will occur. Once its parameters have been determined, a model can be judged by how well it removes autocorrelation from the u_i^2 .

The GARCH(1,1) model can be used to estimate a volatility for options from historical data. This analysis is often used to calculate the impact of a shock to volatility on the implied volatilities of options of different maturities.

Further Reading

On the Causes of Volatility

- Fama, E. F. "The Behavior of Stock Market Prices." *Journal of Business* 38 (January 1965): 34–105.
- French, K. R. "Stock Returns and the Weekend Effect." *Journal of Financial Economics* 8 (March 1980): 55–69.
- French, K. R., and R. Roll. "Stock Return Variances: The Arrival of Information and the Reaction of Traders." *Journal of Financial Economics* 17 (September 1986): 5–26.
- Roll, R. "Orange Juice and Weather." *American Economic Review* 74, no. 5 (December 1984): 861–880.

On GARCH

- Bollerslev, T. "Generalized Autoregressive Conditional Heteroscedasticity." *Journal of Econometrics* 31 (1986): 307–327.
- Cumby, R., S. Figlewski, and J. Hasbrook. "Forecasting Volatilities and Correlations with EGARCH Models." *Journal of Derivatives* 1, no. 2 (Winter 1993): 51–63.
- Engle, R. F. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation." *Econometrica* 50 (1982): 987–1008.
- Engle, R. F., and J. Mezrich. "Grappling with GARCH." *Risk* (September 1995): 112–117.
- Engle, R. F., and V. Ng. "Measuring and Testing the Impact of News on Volatility." *Journal of Finance* 48 (1993): 1749–1778.
- Nelson, D. "Conditional Heteroscedasticity and Asset Returns; A New Approach." *Econometrica* 59 (1990): 347–370.
- Noh, J., R. F. Engle, and A. Kane. "Forecasting Volatility and Option Prices of the S&P 500 Index." *Journal of Derivatives* 2 (1994): 17–30.

Practice Questions and Problems (Answers at End of Book)

- 8.1 The volatility of an asset is 2% per day. What is the standard deviation of the percentage price change in three days?
- 8.2 The volatility of an asset is 25% per annum. What is the standard deviation of the percentage price change in one trading day? Assuming a normal distribution with zero mean, estimate 95% confidence limits for the percentage price change in one day.
- 8.3 Why do traders assume 252 rather than 365 days in a year when using volatilities?
- 8.4 What is implied volatility? What does it mean if different options on the same asset have different implied volatilities?
- 8.5 Suppose that observations on an exchange rate at the end of the past 11 days have been 0.7000, 0.7010, 0.7070, 0.6999, 0.6970, 0.7003, 0.6951, 0.6953, 0.6934, 0.6923, and 0.6922. Estimate the daily volatility using both equation (8.2) and equation (8.4).
- 8.6 The number of visitors to websites follows the power law in equation (8.1) with $\alpha = 2$. Suppose that 1% of sites get 500 or more visitors per day. What percentage of sites get (a) 1,000 and (b) 2,000 or more visitors per day?
- 8.7 Explain the exponentially weighted moving average (EWMA) model for estimating volatility from historical data.
- 8.8 What is the difference between the exponentially weighted moving average model and the GARCH(1,1) model for updating volatilities?
- 8.9 The most recent estimate of the daily volatility of an asset is 1.5% and the price of the asset at the close of trading yesterday was \$30.00. The parameter λ in the EWMA model is 0.94. Suppose that the price of the asset at the close of trading today is \$30.50. How will this cause the volatility to be updated by the EWMA model?

- 8.10 A company uses an EWMA model for forecasting volatility. It decides to change the parameter λ from 0.95 to 0.85. Explain the likely impact on the forecasts.
- 8.11 Assume that an index at close of trading yesterday was 1,040 and the daily volatility of the index was estimated as 1% per day at that time. The parameters in a GARCH(1,1) model are $\omega = 0.000002$, $\alpha = 0.06$, and $\beta = 0.92$. If the level of the index at close of trading today is 1,060, what is the new volatility estimate?
- 8.12 The most recent estimate of the daily volatility of the dollar–sterling exchange rate is 0.6% and the exchange rate at 4:00 P.M. yesterday was 1.5000. The parameter λ in the EWMA model is 0.9. Suppose that the exchange rate at 4:00 P.M. today proves to be 1.4950. How would the estimate of the daily volatility be updated?
- 8.13 A company uses the GARCH(1,1) model for updating volatility. The three parameters are ω , α , and β . Describe the impact of making a small increase in each of the parameters while keeping the others fixed.
- 8.14 The parameters of a GARCH(1,1) model are estimated as $\omega = 0.000004$, $\alpha = 0.05$, and $\beta = 0.92$. What is the long-run average volatility and what is the equation describing the way that the variance rate reverts to its long-run average? If the current volatility is 20% per year, what is the expected volatility in 20 days?
- 8.15 Suppose that the daily volatility of the FTSE 100 stock index (measured in pounds sterling) is 1.8% and the daily volatility of the dollar–sterling exchange rate is 0.9%. Suppose further that the correlation between the FTSE 100 and the dollar–sterling exchange rate is 0.4. What is the volatility of the FTSE 100 when it is translated to U.S. dollars? Assume that the dollar–sterling exchange rate is expressed as the number of U.S. dollars per pound sterling. (*Hint: When $Z = XY$, the percentage daily change in Z is approximately equal to the percentage daily change in X plus the percentage daily change in Y .*)
- 8.16 Suppose that GARCH(1,1) parameters have been estimated as $\omega = 0.000003$, $\alpha = 0.04$, and $\beta = 0.94$. The current daily volatility is estimated to be 1%. Estimate the daily volatility in 30 days.
- 8.17 Suppose that GARCH(1,1) parameters have been estimated as $\omega = 0.000002$, $\alpha = 0.04$, and $\beta = 0.94$. The current daily volatility is estimated to be 1.3%. Estimate the volatility per annum that should be used to price a 20-day option.

Further Questions

- 8.18 Suppose that observations on a stock price (in dollars) at the end of each of 15 consecutive days are as follows:
- 30.2, 32.0, 31.1, 30.1, 30.2, 30.3, 30.6, 30.9, 30.5, 31.1, 31.3, 30.8, 30.3, 29.9, 29.8
- Estimate the daily volatility using both equation (8.2) and equation (8.4).
- 8.19 Suppose that the price of an asset at close of trading yesterday was \$300 and its volatility was estimated as 1.3% per day. The price at the close of trading today is \$298. Update the volatility estimate using

- (a) The EWMA model with $\lambda = 0.94$
 - (b) The GARCH(1,1) model with $\omega = 0.000002$, $\alpha = 0.04$, and $\beta = 0.94$.
- 8.20 An Excel spreadsheet containing over 900 days of daily data on a number of different exchange rates and stock indices can be downloaded from the author's website: www-2.rotman.utoronto.ca/~hull/data. Choose one exchange rate and one stock index. Estimate the value of λ in the EWMA model that minimizes the value of

$$\sum_i (v_i - \beta_i)^2$$

where v_i is the variance forecast made at the end of day $i - 1$ and β_i is the variance calculated from data between day i and day $i + 25$. Use the Solver tool in Excel. To start the EWMA calculations, set the variance forecast at the end of the first day equal to the square of the return on that day.

- 8.21 Suppose that the parameters in a GARCH(1,1) model are $\alpha = 0.03$, $\beta = 0.95$, and $\omega = 0.000002$.
- (a) What is the long-run average volatility?
 - (b) If the current volatility is 1.5% per day, what is your estimate of the volatility in 20, 40, and 60 days?
 - (c) What volatility should be used to price 20-, 40-, and 60-day options?
 - (d) Suppose that there is an event that increases the volatility from 1.5% per day to 2% per day. Estimate the effect on the volatility in 20, 40, and 60 days.
 - (e) Estimate by how much the event increases the volatilities used to price 20-, 40-, and 60-day options.
- 8.22 Estimate parameters for the EWMA and GARCH(1,1) models on the euro–USD exchange rate data between July 27, 2005, and July 27, 2010. These data can be found on the author's website: www-2.rotman.utoronto.ca/~hull/data.
- 8.23 The probability that the loss from a portfolio will be greater than \$10 million in one month is estimated to be 5%.
- (a) What is the loss that has a 1% chance of being exceeded, assuming the change in value of the portfolio is normally distributed with zero mean?
 - (b) What is the loss that has a 1% chance of being exceeded, assuming that the power law applies with $\alpha = 3$?

Chapter 9

Correlations and Copulas

Suppose that a company has an exposure to two different market variables. In the case of each variable, it gains \$10 million if there is a one-standard-deviation increase and loses \$10 million if there is a one-standard-deviation decrease. If changes in the two variables have a high positive correlation, the company's total exposure is very high; if they have a correlation of zero, the exposure is less but still quite large; if they have a high negative correlation, the exposure is quite low because a loss on one of the variables is likely to be offset by a gain on the other. This example shows that it is important for a risk manager to estimate correlations between the changes in market variables as well as their volatilities when assessing risk exposures.

This chapter explains how correlations can be monitored in a similar way to volatilities. It also covers what are known as copulas. These are tools that provide a way of defining a correlation structure between two or more variables, regardless of the shapes of their probability distributions. Copulas have a number of applications in risk management. The chapter shows how a copula can be used to create a model of default correlation for a portfolio of loans. This model is used in the Basel II capital requirements.

9.1 Definition of Correlation

The coefficient of correlation, ρ , between two variables V_1 and V_2 is defined as

$$\rho = \frac{E(V_1 V_2) - E(V_1)E(V_2)}{SD(V_1)SD(V_2)} \quad (9.1)$$

where $E(\cdot)$ denotes expected value and $SD(\cdot)$ denotes standard deviation. If there is no correlation between the variables, $E(V_1 V_2) = E(V_1)E(V_2)$ and $\rho = 0$. If $V_1 = V_2$, both the numerator and the denominator in the expression for ρ equal the variance of V_1 . As we would expect, $\rho = 1$ in this case.

The *covariance* between V_1 and V_2 is defined as

$$\text{cov}(V_1, V_2) = E(V_1 V_2) - E(V_1)E(V_2) \quad (9.2)$$

so that the correlation can be written

$$\rho = \frac{\text{cov}(V_1, V_2)}{SD(V_1)SD(V_2)}$$

Although it is easier to develop intuition about the meaning of a correlation than a covariance, it is covariances that will prove to be the fundamental variables of our analysis. An analogy here is that variance rates were the fundamental variables for the EWMA and GARCH models in Chapter 8, even though it is easier to develop intuition about volatilities.

9.1.1 Correlation vs. Dependence

Two variables are defined as statistically independent if knowledge about one of them does not affect the probability distribution for the other. Formally, V_1 and V_2 are independent if:

$$f(V_2 | V_1 = x) = f(V_2)$$

for all x where $f(\cdot)$ denotes the probability density function and $|$ is the symbol denoting “conditional on.”

If the coefficient of correlation between two variables is zero, does this mean that there is no dependence between the variables? The answer is no. We can illustrate this with a simple example. Suppose that there are three equally likely values for V_1 : -1 , 0 , and $+1$. If $V_1 = -1$ or $V_1 = +1$, then $V_2 = +1$. If $V_1 = 0$, then $V_2 = 0$. In this case, there is clearly a dependence between V_1 and V_2 . If we observe the value of V_1 , we know the value of V_2 . Also, a knowledge of the value of V_2 will cause us to change our probability distribution for V_1 . However, because $E(V_1 V_2) = 0$ and $E(V_1) = 0$, it is easy to see that the coefficient of correlation between V_1 and V_2 is zero.

This example emphasizes the point that the coefficient of correlation measures one particular type of dependence between two variables. This is linear dependence. There are many other ways in which two variables can be related. We can characterize the nature of the dependence between V_1 and V_2 by plotting $E(V_2)$ against V_1 . Three examples are shown in Figure 9.1. Figure 9.1(a) shows linear dependence where the expected value of V_2 depends linearly on V_1 . Figure 9.1(b) shows a V-shaped relationship between the expected value of V_2 and the value of V_1 . (This is similar to the simple example just considered; a symmetrical V-shaped relationship, however strong, leads to

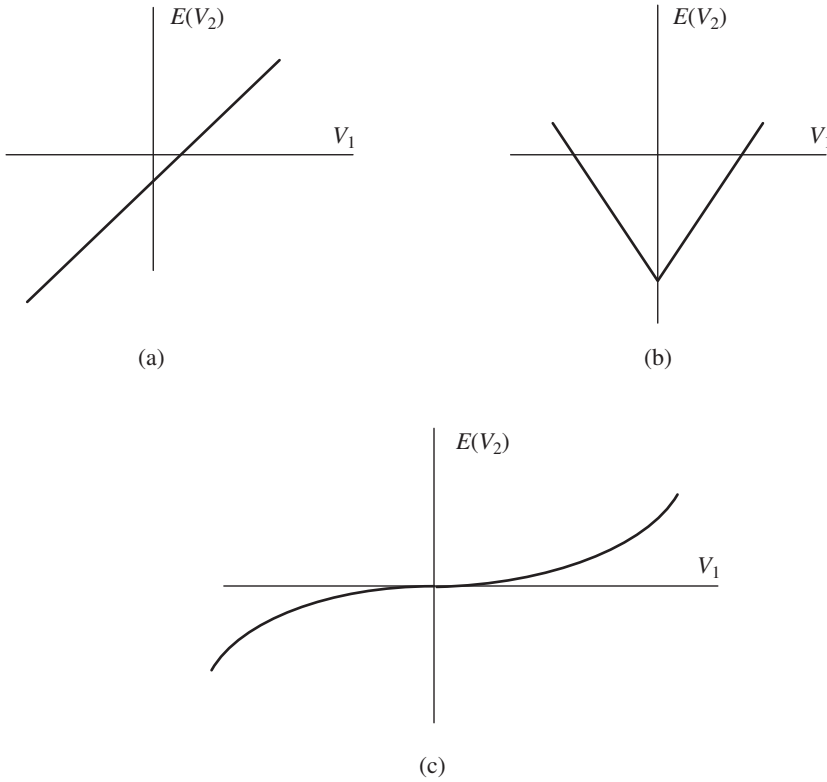


Figure 9.1 Examples of Ways in Which V_2 Can Be Dependent on V_1

zero coefficient of correlation.) Figure 9.1(c) shows a type of dependence that is often seen when V_1 and V_2 are percentage changes in financial variables. For the values of V_1 normally encountered, there is very little relation between V_1 and V_2 . However, extreme values of V_1 tend to lead to extreme values of V_2 . (This could be consistent with correlations increasing in stressed market conditions.)

Another aspect of the way in which V_2 depends on V_1 is found by examining the standard deviation of V_2 conditional on V_1 . As we will see later, this is constant when V_1 and V_2 have a bivariate normal distribution. But, in other situations, the standard deviation of V_2 is liable to depend on the value of V_1 .

9.2 Monitoring Correlation

Chapter 8 explained how exponentially weighted moving average and GARCH models can be developed to monitor the variance rate of a variable. Similar approaches can be used to monitor the covariance rate between two variables. The variance rate per day of a variable is the variance of daily returns. Similarly, the *covariance rate* per day between two variables is defined as the covariance between the daily returns of the variables.

Suppose that X_i and Y_i are the values of two variables, X and Y , at the end of day i . The returns on the variables on day i are

$$x_i = \frac{X_i - X_{i-1}}{X_{i-1}} \quad y_i = \frac{Y_i - Y_{i-1}}{Y_{i-1}}$$

The covariance rate between X and Y on day n is from equation (9.2):

$$\text{cov}_n = E(x_n y_n) - E(x_n)E(y_n)$$

In Section 8.5, we explained that risk managers assume that expected daily returns are zero when the variance rate per day is calculated. They do the same when calculating the covariance rate per day. This means that the covariance rate per day between X and Y on day n is assumed to be

$$\text{cov}_n = E(x_n y_n)$$

Using equal weights for the last m observations on x_i and y_i gives the estimate

$$\text{cov}_n = \frac{1}{m} \sum_{i=1}^m x_{n-i} y_{n-i} \quad (9.3)$$

A similar weighting scheme for variances gives an estimate for the variance rate on day n for variable X as

$$\text{var}_{x,n} = \frac{1}{m} \sum_{i=1}^m x_{n-i}^2$$

and for variable Y as

$$\text{var}_{y,n} = \frac{1}{m} \sum_{i=1}^m y_{n-i}^2$$

The correlation estimate on day n is

$$\frac{\text{cov}_n}{\sqrt{\text{var}_{x,n} \text{var}_{y,n}}}$$

9.2.1 EWMA

Most risk managers would agree that observations from long ago should not have as much weight as recent observations. In Chapter 8, we discussed the use of the exponentially weighted moving average (EWMA) model for variances. We saw that it leads to weights that decline exponentially as we move back through time. A similar weighting scheme can be used for covariances. The formula for updating a covariance estimate in the EWMA model is similar to that in equation (8.8) for variances:

$$\text{cov}_n = \lambda \text{cov}_{n-1} + (1 - \lambda) x_{n-1} y_{n-1}$$

A similar analysis to that presented for the EWMA volatility model shows that the weight given to $x_{n-i}y_{n-i}$ declines as i increases (i.e., as we move back through time). The lower the value of λ , the greater the weight that is given to recent observations.

Example 9.1

Suppose that $\lambda = 0.95$ and that the estimate of the correlation between two variables X and Y on day $n - 1$ is 0.6. Suppose further that the estimates of the volatilities for X and Y on day $n - 1$ are 1% and 2%, respectively. From the relationship between correlation and covariance, the estimate of the covariance rate between X and Y on day $n - 1$ is

$$0.6 \times 0.01 \times 0.02 = 0.00012$$

Suppose that the percentage changes in X and Y on day $n - 1$ are 0.5% and 2.5%, respectively. The variance rates and covariance rate for day n would be updated as follows:

$$\sigma_{x,n}^2 = 0.95 \times 0.01^2 + 0.05 \times 0.005^2 = 0.00009625$$

$$\sigma_{y,n}^2 = 0.95 \times 0.02^2 + 0.05 \times 0.025^2 = 0.00041125$$

$$\text{cov}_n = 0.95 \times 0.00012 + 0.05 \times 0.005 \times 0.025 = 0.00012025$$

The new volatility of X is $\sqrt{0.00009625} = 0.981\%$, and the new volatility of Y is $\sqrt{0.00041125} = 2.028\%$. The new correlation between X and Y is

$$\frac{0.00012025}{0.00981 \times 0.02028} = 0.6044$$

9.2.2 GARCH

GARCH models can also be used for updating covariance rate estimates and forecasting the future level of covariance rates. For example, the GARCH(1,1) model for updating a covariance rate between X and Y is

$$\text{cov}_n = \omega + \alpha x_{n-1}y_{n-1} + \beta \text{cov}_{n-1}$$

This formula, like its counterpart in equation (8.10) for updating variances, gives some weight to a long-run average covariance, some to the most recent covariance estimate, and some to the most recent observation on covariance (which is $x_{n-1}y_{n-1}$). The long-term average covariance rate is $\omega/(1 - \alpha - \beta)$. Formulas similar to those in equations (8.14) and (8.15) can be developed for forecasting future covariance rates and calculating the average covariance rate during a future time period.

9.3 Correlation and Covariance Matrices

Once variances, covariances, and correlations have been calculated for a set of variables, correlation matrices and covariance matrices can be produced.

A correlation matrix is shown in Table 9.1. It displays coefficients of correlation. Because a variable is always perfectly correlated with itself, the diagonal elements of a correlation matrix are always 1. Furthermore, because $\rho_{ij} = \rho_{ji}$, the correlation matrix is symmetric.

It is often more convenient to work with a variance-covariance matrix. When $i \neq j$, the (i, j) element in this matrix is the covariance between variable i and variable j . When $i = j$, it is the variance of variable i . (See Table 9.2.)

9.3.1 Consistency Condition for Covariances

Not all variance-covariance matrices are internally consistent. The condition for an $N \times N$ variance-covariance matrix, Ω , to be internally consistent is

$$\mathbf{w}^T \Omega \mathbf{w} \geq 0 \quad (9.4)$$

for all $N \times 1$ vectors $\mathbf{w}$ where $\mathbf{w}^T$ is the transpose of $\mathbf{w}$. A matrix that satisfies this property is known as *positive-semidefinite*.

To understand why the condition in equation (9.4) must hold, suppose that $\mathbf{w}$ is the (column) vector $(w_1, w_2, \dots, w_N)$. The expression $\mathbf{w}^T \Omega \mathbf{w}$ is the variance rate of a

Table 9.1 A Correlation Matrix

$$\begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \rho_{23} & \dots & \rho_{2n} \\ \rho_{31} & \rho_{32} & 1 & \dots & \rho_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{n1} & \rho_{n2} & \rho_{n3} & \dots & 1 \end{pmatrix}$$

ρ_{ij} is the correlation between variable i and variable j .

Table 9.2 A Variance-Covariance Matrix

$$\begin{pmatrix} \text{var}_1 & \text{cov}_{12} & \text{cov}_{13} & \dots & \text{cov}_{1n} \\ \text{cov}_{21} & \text{var}_2 & \text{cov}_{23} & \dots & \text{cov}_{2n} \\ \text{cov}_{31} & \text{cov}_{32} & \text{var}_3 & \dots & \text{cov}_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \text{cov}_{n1} & \text{cov}_{n2} & \text{cov}_{n3} & \dots & \text{var}_n \end{pmatrix}$$

cov_{ij} is the covariance between variable i and variable j and $\text{var}_i = \text{cov}_{ii}$ is the variance of variable i .

portfolio where an amount w_i is invested in market variable i . As such, it cannot be negative.

To ensure that a positive-semidefinite matrix is produced, variances and covariances should be calculated consistently. For example, if variance rates are calculated by giving equal weight to the last m data items, the same should be done for covariance rates. If variance rates are updated using an EWMA model with $\lambda = 0.94$, the same should be done for covariance rates. Using a GARCH model to update a variance-covariance matrix in a consistent way is trickier and requires a multivariate model.¹

An example of a variance-covariance matrix that is not internally consistent is

$$\begin{pmatrix} 1 & 0 & 0.9 \\ 0 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \end{pmatrix}$$

The variance of each variable is 1.0 and so the covariances are also coefficients of correlation in this case. The first variable is highly correlated with the third variable, and the second variable is also highly correlated with the third variable. However, there is no correlation at all between the first and second variables. This seems strange. When we set $\mathbf{w}^T$ equal to $(1, 1, -1)$, we find that the condition in equation (9.4) is not satisfied, proving that the matrix is not positive-semidefinite.²

If we make a small change to a positive-semidefinite matrix that is calculated from observations on three variables (e.g., for the purposes of doing a sensitivity analysis), it is likely that the matrix will remain positive-semidefinite. However, if we do the same thing for observations on 100 variables, we have to be much more careful. An arbitrary small change to a positive-semidefinite 100×100 matrix is quite likely to lead to it no longer being positive-semidefinite.

9.4 Multivariate Normal Distributions

Multivariate normal distributions are well understood and relatively easy to deal with. As we will explain in the next section, they can be useful tools for specifying the correlation structure between variables, even when the distributions of the variables are not normal.

We start by considering a bivariate normal distribution where there are only two variables, V_1 and V_2 . Suppose that we know that V_1 has some value. Conditional on this, the value of V_2 is normal with mean

$$\mu_2 + \rho\sigma_2 \frac{V_1 - \mu_1}{\sigma_1}$$

¹See R. Engle and J. Mezrich, “GARCH for Groups,” *Risk* (August 1996): 36–40, for a discussion of alternative approaches.

²It can be shown that the condition for a 3×3 matrix of correlations to be internally consistent is

$$\rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2 - 2\rho_{12}\rho_{13}\rho_{23} \leq 1$$

where ρ_{ij} is the coefficient of correlation between variables i and j .

and standard deviation

$$\sigma_2 \sqrt{1 - \rho^2}$$

Here μ_1 and μ_2 are the unconditional means of V_1 and V_2 , σ_1 and σ_2 are their unconditional standard deviations, and ρ is the coefficient of correlation between V_1 and V_2 . Note that the expected value of V_2 conditional on V_1 is linearly dependent on the value of V_1 . This corresponds to Figure 9.1(a). Also, the standard deviation of V_2 conditional on the value of V_1 is the same for all values of V_1 .

Software for calculating the cumulative bivariate normal distribution is on the author's website:

www-2.rotman.utoronto.ca/~hull/riskman

9.4.1 Generating Random Samples from Multivariate Normal Distributions

Most programming languages have routines for sampling a random number between zero and one, and many have routines for sampling from a normal distribution.³

When samples ε_1 and ε_2 from a bivariate normal distribution (where both variables have mean zero and standard deviation one) are required, the usual procedure involves first obtaining independent samples z_1 and z_2 from a univariate standard normal distribution (i.e., a normal distribution with mean zero and standard deviation one). The required samples ε_1 and ε_2 are then calculated as follows:

$$\begin{aligned}\varepsilon_1 &= z_1 \\ \varepsilon_2 &= \rho z_1 + z_2 \sqrt{1 - \rho^2}\end{aligned}$$

where ρ is the coefficient of correlation in the bivariate normal distribution.

Consider next a situation where we require samples from a multivariate normal distribution (where all variables have mean zero and standard deviation one) and the coefficient of correlation between variable i and variable j is ρ_{ij} . We first sample n independent variables z_i ($1 \leq i \leq n$) from univariate standard normal distributions. The required samples are ε_i ($1 \leq i \leq n$), where

$$\varepsilon_i = \sum_{k=1}^i \alpha_{ik} z_k \tag{9.5}$$

and the α_{ik} are parameters chosen to give the correct variances and the correct correlations for ε_i . For $1 \leq j < i$, we must have

$$\sum_{k=1}^i \alpha_{ik}^2 = 1$$

³In Excel, the instruction =NORMSINV(RAND()) gives a random sample from a normal distribution.

and, for all $j < i$,

$$\sum_{k=1}^j \alpha_{ik} \alpha_{jk} = \rho_{ij}$$

The first sample, ε_1 , is set equal to z_1 . These equations can be solved so that ε_2 is calculated from z_1 and z_2 , ε_3 is calculated from z_1 , z_2 , and z_3 , and so on. The procedure is known as the *Cholesky decomposition*. (See Problem 9.9.)

If we find ourselves trying to take the square root of a negative number when using the Cholesky decomposition, the variance-covariance matrix assumed for the variables is not internally consistent. As explained in Section 9.3.1, this is equivalent to saying that the matrix is not positive-semidefinite.

9.4.2 Factor Models

Sometimes the correlations between normally distributed variables are defined using a factor model. Suppose that $U_1, U_2, \dots, U_N$ have standard normal distributions (i.e., normal distributions with mean zero and standard deviation one). In a one-factor model, each U_i ($1 \leq i \leq N$) has a component dependent on a common factor, F , and a component that is uncorrelated with the other variables. Formally,

$$U_i = a_i F + \sqrt{1 - a_i^2} Z_i \quad (9.6)$$

where F and the Z_i have standard normal distributions and a_i is a constant between -1 and $+1$. The Z_i are uncorrelated with each other and uncorrelated with F . The coefficient of Z_i is chosen so that U_i has a mean of zero and a variance of one. In this model, all the correlation between U_i and U_j arises from their dependence on the common factor, F . The coefficient of correlation between U_i and U_j is $a_i a_j$.

A one-factor model imposes some structure on the correlations and has the advantage that the resulting covariance matrix is always positive-semidefinite. Without assuming a factor model, the number of correlations that have to be estimated for the N variables is $N(N-1)/2$. With the one-factor model, we need only estimate N parameters: $a_1, a_2, \dots, a_N$. An example of a one-factor model from the world of investments is the capital asset pricing model where the return on a stock has a component dependent on the return from the market and an idiosyncratic (nonsystematic) component that is independent of the return on other stocks (see Section 1.3).

The one-factor model can be extended to a two-factor, three-factor, or M -factor model. In the M -factor model

$$U_i = a_{i1} F_1 + a_{i2} F_2 + \dots + a_{iM} F_M + \sqrt{1 - a_{i1}^2 - a_{i2}^2 - \dots - a_{iM}^2} Z_i \quad (9.7)$$

The factors, $F_1, F_2, \dots, F_M$ have uncorrelated standard normal distributions and the Z_i are uncorrelated both with each other and with the factors. In this case, the correlation between U_i and U_j is

$$\sum_{m=1}^M a_{im} a_{jm}$$

9.5 Copulas

Consider two correlated variables, V_1 and V_2 . The *marginal distribution* of V_1 (sometimes also referred to as the unconditional distribution) is its distribution assuming we know nothing about V_2 ; similarly, the marginal distribution of V_2 is its distribution assuming we know nothing about V_1 . Suppose we have estimated the marginal distributions of V_1 and V_2 . How can we make an assumption about the correlation structure between the two variables to define their joint distribution?

If the marginal distributions of V_1 and V_2 are normal, a convenient and easy-to-work-with assumption is that the joint distribution of the variables is bivariate normal.⁴ (The correlation structure between the variables is then as described in Section 9.4.) Similar assumptions are possible for some other marginal distributions. But often there is no natural way of defining a correlation structure between two marginal distributions. This is where copulas come in.

As an example of the application of copulas, suppose that variables V_1 and V_2 have the triangular probability density functions shown in Figure 9.2. Both variables have values between 0 and 1. The density function for V_1 peaks at 0.2. The density function for V_2 peaks at 0.5. For both density functions, the maximum height is 2.0 (so that the area under the density function is 1.0). To use what is known as a *Gaussian copula*, we map V_1 and V_2 into new variables U_1 and U_2 that have standard normal distributions. (A standard normal distribution is a normal distribution with mean zero and standard deviation one.) The mapping is accomplished on a percentile-to-percentile basis. The one-percentile point of the V_1 distribution is mapped to the one-percentile point of the U_1 distribution; the 10-percentile point of the V_1 distribution is mapped to the

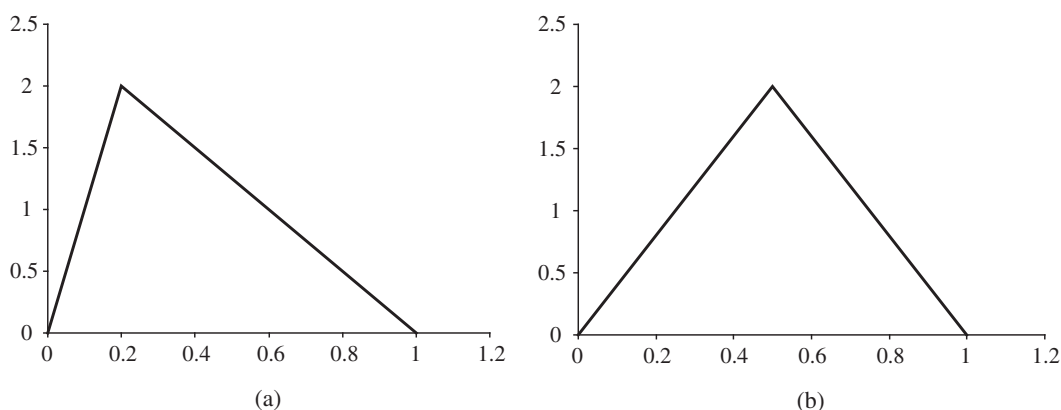


Figure 9.2 Triangular Distributions for V_1 and V_2

⁴Although the bivariate normal assumption is a convenient one, it is not the only one that can be made. There are many other ways in which two normally distributed variables can be dependent on each other. For example, we could have $V_2 = V_1$ for $-k \leq V_1 \leq k$ and $V_2 = -V_1$ otherwise. See also Problem 9.11.

Table 9.3 Mapping of V_1 That Has the Triangular Distribution in Figure 9.2(a) to U_1 That Has a Standard Normal Distribution

V_1 Value	Percentile of Distribution	U_1 Value
0.1	5.00	-1.64
0.2	20.00	-0.84
0.3	38.75	-0.29
0.4	55.00	0.13
0.5	68.75	0.49
0.6	80.00	0.84
0.7	88.75	1.21
0.8	95.00	1.64
0.9	98.75	2.24

Table 9.4 Mapping of V_2 That Has the Triangular Distribution in Figure 9.2(b) to U_2 That Has a Standard Normal Distribution

V_2 Value	Percentile of Distribution	U_2 Value
0.1	2.00	-2.05
0.2	8.00	-1.41
0.3	18.00	-0.92
0.4	32.00	-0.47
0.5	50.00	0.00
0.6	68.00	0.47
0.7	82.00	0.92
0.8	92.00	1.41
0.9	98.00	2.05

10-percentile point of the U_1 distribution; and so on. V_2 is mapped into U_2 in a similar way. Table 9.3 shows how values of V_1 are mapped into values of U_1 . Table 9.4 similarly shows how values of V_2 are mapped into values of U_2 . Consider the $V_1 = 0.1$ calculation in Table 9.3. The cumulative probability that V_1 is less than 0.1 is (by calculating areas of triangles) $0.5 \times 0.1 \times 1 = 0.05$ or 5%. The value 0.1 for V_1 therefore gets mapped to the five-percentile point of the standard normal distribution. This is -1.64 .⁵

The variables, U_1 and U_2 , have normal distributions. We assume that they are jointly bivariate normal. This in turn implies a joint distribution and a correlation structure between V_1 and V_2 . The essence of the copula is therefore that, instead of defining a correlation structure between V_1 and V_2 directly, we do so indirectly. We map V_1 and

⁵It can be calculated using Excel: $\text{NORMSINV}(0.05) = -1.64$.

V_2 into other variables that have well-behaved distributions and for which it is easy to define a correlation structure.

Suppose that we assume that the correlation between U_1 and U_2 is 0.5. The joint cumulative probability distribution between V_1 and V_2 is shown in Table 9.5. To illustrate the calculations, consider the first one where we are calculating the probability that $V_1 < 0.1$ and $V_2 < 0.1$. From Tables 9.3 and 9.4, this is the same as the probability that $U_1 < -1.64$ and $U_2 < -2.05$. From the cumulative bivariate normal distribution, this is 0.006 when $\rho = 0.5$.⁶ (Note that the probability would be only $0.02 \times 0.05 = 0.001$ if $\rho = 0$.)

The correlation between U_1 and U_2 is referred to as the *copula correlation*. This is not, in general, the same as the coefficient of correlation between V_1 and V_2 . Because U_1 and U_2 are bivariate normal, the conditional mean of U_2 is linearly dependent on U_1 and the conditional standard deviation of U_2 is constant (as discussed in Section 9.4). However, a similar result does not in general apply to V_1 and V_2 .

9.5.1 Expressing the Approach Algebraically

The way in which a Gaussian copula defines a joint distribution is illustrated in Figure 9.3. For a more formal description of the model, suppose that G_1 and G_2 are the cumulative marginal (i.e., unconditional) probability distributions of V_1 and V_2 . We map $V_1 = v_1$ to $U_1 = u_1$ and $V_2 = v_2$ to $U_2 = u_2$ so that

$$G_1(v_1) = N(u_1)$$

and

$$G_2(v_2) = N(u_2)$$

Table 9.5 Cumulative Joint Probability Distribution for V_1 and V_2 in the Gaussian Copula Model (Correlation parameter = 0.5. Table shows the joint probability that V_1 and V_2 are less than the specified values.)

V_1	V_2								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.006	0.017	0.028	0.037	0.044	0.048	0.049	0.050	0.050
0.2	0.013	0.043	0.081	0.120	0.156	0.181	0.193	0.198	0.200
0.3	0.017	0.061	0.124	0.197	0.273	0.331	0.364	0.381	0.387
0.4	0.019	0.071	0.149	0.248	0.358	0.449	0.505	0.535	0.548
0.5	0.019	0.076	0.164	0.281	0.417	0.537	0.616	0.663	0.683
0.6	0.020	0.078	0.173	0.301	0.456	0.600	0.701	0.763	0.793
0.7	0.020	0.079	0.177	0.312	0.481	0.642	0.760	0.837	0.877
0.8	0.020	0.080	0.179	0.318	0.494	0.667	0.798	0.887	0.936
0.9	0.020	0.080	0.180	0.320	0.499	0.678	0.816	0.913	0.970

⁶An Excel function for calculating the cumulative bivariate normal distribution is on the author's website: www-2.rotman.utoronto.ca/~hull/riskman.

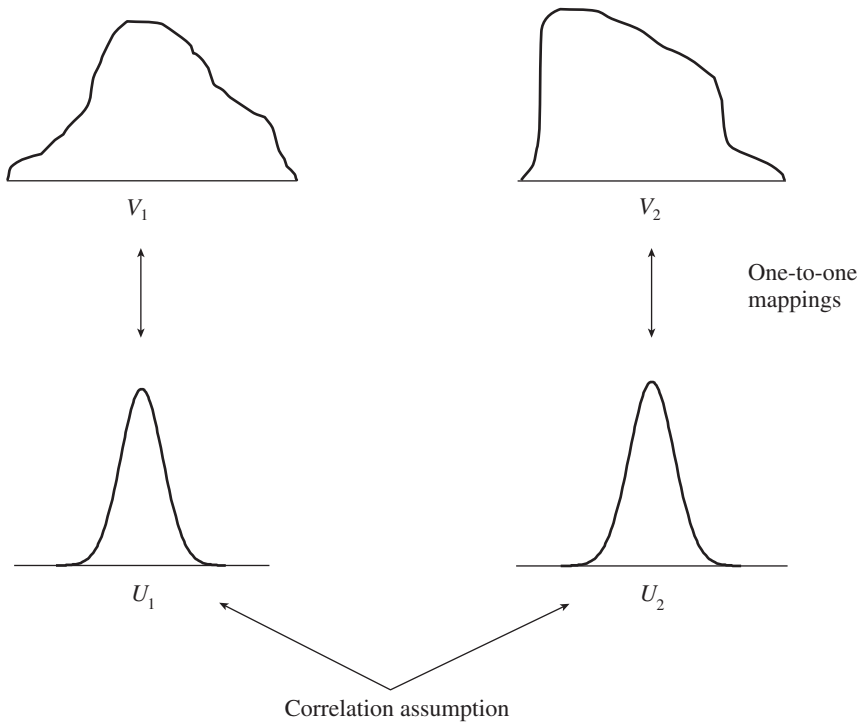


Figure 9.3 The Way in Which a Copula Model Defines a Joint Distribution

where N is the cumulative normal distribution function. This means that

$$\begin{aligned} u_1 &= N^{-1}[G_1(v_1)] & u_2 &= N^{-1}[G_2(v_2)] \\ v_1 &= G_1^{-1}[N(u_1)] & v_2 &= G_2^{-1}[N(u_2)] \end{aligned}$$

The variables U_1 and U_2 are then assumed to be bivariate normal. The key property of a copula model is that it preserves the marginal distributions of V_1 and V_2 (however unusual these may be) while defining a correlation structure between them.

9.5.2 Other Copulas

The Gaussian copula is just one copula that can be used to define a correlation structure between V_1 and V_2 . There are many other copulas leading to many other correlation structures. One that is sometimes used is the *Student's t-copula*. This works in the same way as the Gaussian copula except that the variables U_1 and U_2 are assumed to have a bivariate Student's t -distribution instead of a bivariate normal distribution. To sample from a bivariate Student's t -distribution with f degrees of freedom and correlation ρ , the steps are as follows:

1. Sample from the inverse chi-square distribution to get a value χ . (In Excel, the CHI-INV function can be used. The first argument is RAND() and the second is f .)

2. Sample from a bivariate normal distribution with correlation ρ as described in Section 9.4.
3. Multiply the normally distributed samples by $\sqrt{f/\chi}$.

9.5.3 Tail Dependence

Figure 9.4 shows plots of 5,000 random samples from a bivariate normal distribution, while Figure 9.5 does the same for the bivariate Student's t . The correlation parameter is 0.5 and the number of degrees of freedom for the Student's t is 4. Define a tail value of a distribution as a value in the left or right 1% tail of the distribution. There is a tail value for the normal distribution when the variable is greater than 2.33 or less than -2.33 . Similarly, there is a tail value in the Student's t -distribution when the value of the variable is greater than 3.75 or less than -3.75 . Vertical and horizontal lines in the figures indicate when tail values occur. The figures illustrate that it is much more common for the two variables to have tail values at the same time in the bivariate Student's t -distribution than in the bivariate normal distribution. To put this another way, the *tail dependence* is higher in a bivariate Student's t -distribution than in a bivariate normal distribution. We made the point earlier that correlations between market variables tend to increase in extreme market conditions, so that Figure 9.1(c) is sometimes a better description of the correlation structure between two variables than Figure 9.1(a). This has led some researchers to argue that the Student's t -copula provides a better description of the joint behavior of two market variables than the Gaussian copula.

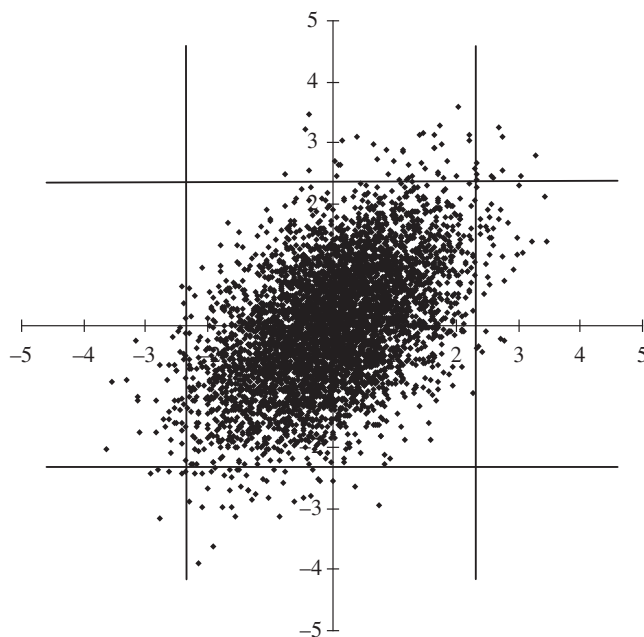


Figure 9.4 5,000 Random Samples from a Bivariate Normal Distribution

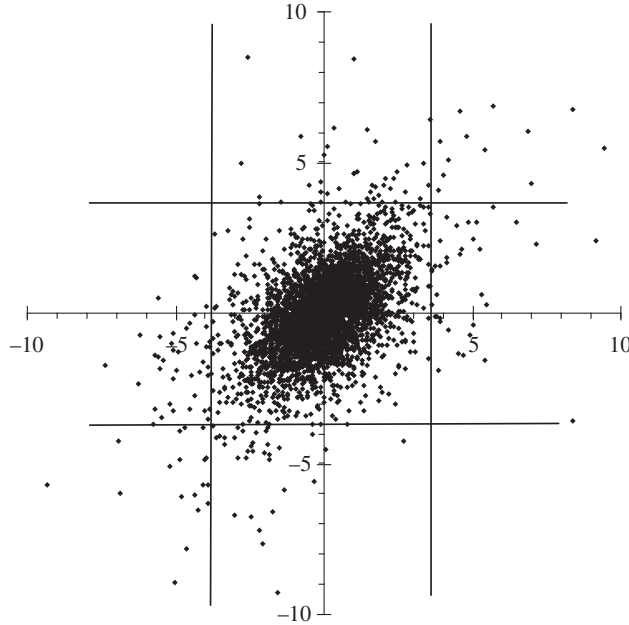


Figure 9.5 5,000 Random Samples from a Bivariate Student's t -distribution with Four Degrees of Freedom

9.5.4 Multivariate Copulas

Copulas can be used to define a correlation structure between more than two variables. The simplest example of this is the multivariate Gaussian copula. Suppose that there are N variables, $V_1, V_2, \dots, V_N$ and that we know the marginal distribution of each variable. For each i ($1 \leq i \leq N$), we transform V_i into U_i where U_i has a standard normal distribution. (As described earlier, the transformation is accomplished on a percentile-to-percentile basis.) We then assume that the U_i have a multivariate normal distribution.

9.5.5 A Factor Copula Model

In multivariate copula models, analysts often assume a factor model for the correlation structure between the U_i . When there is only one factor, equation (9.6) gives

$$U_i = a_i F + \sqrt{1 - a_i^2} Z_i \quad (9.8)$$

where F and the Z_i have standard normal distributions. The Z_i are uncorrelated with each other and with F .

Other factor copula models are obtained by choosing F and the Z_i to have other zero-mean unit-variance distributions. We discuss this further in connection with credit risk in the next section. These distributional choices affect the nature of the dependence between the U -variables and therefore that between the V -variables.

9.6 Application to Loan Portfolios: Vasicek’s Model

We now present an application of the one-factor Gaussian copula model that will prove useful in understanding the Basel II capital requirements, which are discussed in Chapter 25. Suppose a bank has a large portfolio of loans where the probability of default per year for each loan is 1%. If the loans default independently of each other, we would expect the default rate to be almost exactly 1% every year. In practice, loans do not default independently of each other. They are all influenced by macroeconomic conditions. As a result, in some years the default rate is high whereas in others it is low. This is illustrated by Table 9.6, which shows the default rates compiled by S&P for all rated companies between 1981 and 2020. The default rate varies from a low of 0.15% in 1981 to a high of 4.19% in 2009. Other high-default-rate years were 1990, 1991, 1999, 2000, 2001, 2002, 2016, and 2020

To model the defaults of the loans in a portfolio, we define T_i as the time when company i defaults. (There is an implicit assumption that all companies will default eventually—but the default may happen a long time, perhaps even hundreds of years, in the future.) We make the simplifying assumption that all loans have the same unconditional cumulative probability distribution for the time to default and define PD as the probability of default by time T :

$$PD = \text{Prob}(T_i < T).$$

The Gaussian copula model can be used to define a correlation structure between the times to default of the loans. Following the procedure we have described, each time to default T_i is mapped to a variable U_i that has a standard normal distribution on a percentile-to-percentile basis.

Table 9.6 Annual Percentage Default Rate for All Rated Companies, 1981–2020

Year	Default Rate	Year	Default Rate	Year	Default Rate
1981	0.15	1995	1.05	2009	4.19
1982	1.22	1996	0.51	2010	1.21
1983	0.77	1997	0.63	2011	0.80
1984	0.93	1998	1.28	2012	1.14
1985	1.13	1999	2.15	2013	1.06
1986	1.74	2000	2.48	2014	0.69
1987	0.95	2001	3.79	2015	1.36
1988	1.39	2002	3.60	2016	2.09
1989	1.79	2003	1.93	2017	1.21
1990	2.74	2004	0.78	2018	1.03
1991	3.26	2005	0.60	2019	1.30
1992	1.50	2006	0.48	2020	2.74
1993	0.60	2007	0.37		
1994	0.63	2008	1.80		

SOURCE: S&P Global Ratings: www.spglobal.com/ratingsdirect.

We assume the factor model in equation (9.8) for the correlation structure is between the U_i and make the simplifying assumption that the a_i are all the same and equal to a so that:

$$U_i = aF + \sqrt{1 - a^2}Z_i$$

As in equation (9.8), the variables F and Z_i have independent standard normal distributions. The copula correlation between each pair of loans is in this case the same. It is

$$\rho = a^2$$

so that the expression for U_i can be written

$$U_i = \sqrt{\rho}F + \sqrt{1 - \rho}Z_i \quad (9.9)$$

Define the “worst case default rate,” $\text{WCDR}(T, X)$, as the default rate (i.e., percentage of loans defaulting) during time T that will not be exceeded with probability X . (In many applications T will be one year.) As shown in what follows, the assumptions we have made lead to

$$\text{WCDR}(T, X) = N\left(\frac{N^{-1}(\text{PD}) + \sqrt{\rho}N^{-1}(X)}{\sqrt{1 - \rho}}\right) \quad (9.10)$$

This is a strange-looking result, but a very important one. It was first developed by Vasicek in 1987.⁷ N and N^{-1} are the cumulative normal and inverse cumulative normal distribution functions. They can be calculated using the NORMSDIST and NORMSINV functions in Excel. Note that if $\rho = 0$, the loans default independently of each other and $\text{WCDR} = \text{PD}$. As ρ increases, WCDR increases.

Example 9.2

Suppose that a bank has a large number of loans to retail customers. The one-year probability of default for each loan is 2% and the copula correlation parameter, ρ , in Vasicek's model is estimated as 0.1. In this case,

$$\text{WCDR}(1, 0.999) = N\left(\frac{N^{-1}(0.02) + \sqrt{0.1}N^{-1}(0.999)}{\sqrt{1 - 0.1}}\right) = 0.128$$

showing that the 99.9% worst case one-year default rate is 12.8%.

⁷See O. Vasicek, “Probability of Loss on a Loan Portfolio” (Working Paper, KMV, 1987). Vasicek's results were published in *Risk* in December 2002 under the title “Loan Portfolio Value.”

9.6.1 Proof of Vasicek's Result

From the properties of the Gaussian copula model

$$PD = \text{Prob}(T_i < T) = \text{Prob}(U_i < U)$$

where

$$U = N^{-1}[PD] \quad (9.11)$$

The probability of default by time T depends on the value of the factor, F , in equation (9.9). The factor can be thought of as an index of macroeconomic conditions. If F is high, macroeconomic conditions are good. Each U_i will then tend to be high and the corresponding T_i will therefore also tend to be high, meaning that the probability of an early default is low and therefore $\text{Prob}(T_i < T)$ is low. If F is low, macroeconomic conditions are bad. Each U_i and the corresponding T_i will then tend to be low so that the probability of an early default is high. To explore this further, we consider the probability of default conditional on F .

From equation (9.9),

$$Z_i = \frac{U_i - \sqrt{\rho}F}{\sqrt{1 - \rho}}$$

The probability that $U_i < U$ conditional on the factor value, F , is

$$\text{Prob}(U_i < U|F) = \text{Prob}\left(Z_i < \frac{U - \sqrt{\rho}F}{\sqrt{1 - \rho}}\right) = N\left(\frac{U - \sqrt{\rho}F}{\sqrt{1 - \rho}}\right)$$

This is also $\text{Prob}(T_i < T|F)$ so that

$$\text{Prob}(T_i < T|F) = N\left(\frac{U - \sqrt{\rho}F}{\sqrt{1 - \rho}}\right) \quad (9.12)$$

From equation (9.11) this becomes

$$\text{Prob}(T_i < T|F) = N\left(\frac{N^{-1}(PD) - \sqrt{\rho}F}{\sqrt{1 - \rho}}\right) \quad (9.13)$$

For a large portfolio of loans with the same PD, where the copula correlation for each pair of loans is ρ , this equation provides a good estimate of the percentage of loans defaulting by time T conditional on F . We will refer to this as the default rate.

As F decreases, the default rate increases. How bad can the default rate become? Because F has a normal distribution, the probability that F will be less than $N^{-1}(Y)$ is Y . There is therefore a probability of Y that the default rate will be greater than

$$N\left(\frac{N^{-1}(PD) - \sqrt{\rho}N^{-1}(Y)}{\sqrt{1 - \rho}}\right)$$

The default rate that we are $X\%$ certain will not be exceeded in time T is obtained by substituting $Y = 1 - X$ into the preceding expression. Because $N^{-1}(X) = -N^{-1}(1 - X)$, we obtain equation (9.10).

9.6.2 Estimating PD and ρ

The maximum likelihood methods explained in Chapter 8 can be used to estimate PD and ρ from historical data on default rates. We used equation (9.10) to calculate a high percentile of the default rate distribution, but it is actually true for all percentiles. If DR is the default rate and $G(\text{DR})$ is the cumulative probability distribution function for DR, equation (9.10) shows that

$$\text{DR} = N\left(\frac{N^{-1}(\text{PD}) + \sqrt{\rho}N^{-1}(G(\text{DR}))}{\sqrt{1 - \rho}}\right)$$

Rearranging this equation,

$$G(\text{DR}) = N\left(\frac{\sqrt{1 - \rho}N^{-1}(\text{DR}) - N^{-1}(\text{PD})}{\sqrt{\rho}}\right) \quad (9.14)$$

Differentiating this, the probability density function for the default rate is

$$g(\text{DR}) = \sqrt{\frac{1 - \rho}{\rho}} \exp\left\{\frac{1}{2}\left[(N^{-1}(\text{DR}))^2 - \left(\frac{\sqrt{1 - \rho}N^{-1}(\text{DR}) - N^{-1}(\text{PD})}{\sqrt{\rho}}\right)^2\right]\right\} \quad (9.15)$$

The procedure for calculating maximum likelihood estimates for PD and ρ from historical data is as follows:

1. Choose trial values for PD and ρ .
2. Calculate the logarithm of the probability density in equation (9.15) for each of the observations on DR.
3. Use Solver to search for the values of PD and ρ that maximize the sum of the values in 2.

One application of this is to the data in Table 9.6. The estimates for ρ and PD given by these data are 0.063 and 1.48%, respectively. (See worksheet on the author's website for the calculations.) The probability distribution for the default rate is shown in Figure 9.6. The 99.9% worst case default rate is

$$N\left(\frac{N^{-1}(0.0148) + \sqrt{0.063}N^{-1}(0.999)}{\sqrt{1 - 0.063}}\right) = 0.074$$

or 7.4% per annum.

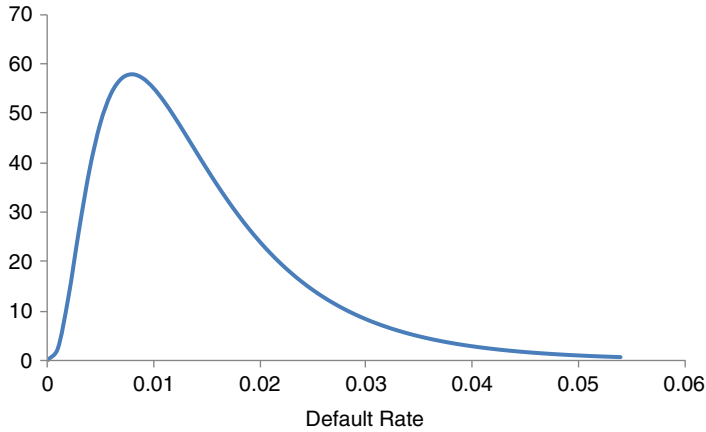


Figure 9.6 Probability Distribution of Default Rate When Parameters Are Estimated Using the Data in Table 9.6

9.6.3 Alternatives to the Gaussian Copula

The one-factor Gaussian copula model has its limitations. As Figure 9.4 illustrates, it leads to very little tail dependence. This means that an unusually early default for one company does not often happen at the same time as an unusually early default for another company. It can be difficult to find a ρ to fit data. For example, there is no ρ that is consistent with a PD of 1% and the situation where one year in 10 the default rate is greater than 3%. Other one-factor copula models with more tail dependence can provide a better fit to data.

An approach to developing other one-factor copulas is to choose F or Z_i , or both, as distributions with heavier tails than the normal distribution in equation (9.9). (They have to be scaled so that they have a mean of zero and standard deviation of one.) The distribution of U_i is then determined (possibly numerically) from the distributions of F and Z_i . Equation (9.10) becomes

$$\text{WCDR}(T, X) = \Phi \left(\frac{\Psi^{-1}(\text{PD}) + \sqrt{\rho} \Theta^{-1}(X)}{\sqrt{1 - \rho}} \right)$$

where Φ , Θ , and Ψ are the cumulative probability distributions of Z_i , F , and U_i , and equation (9.14) becomes⁸

$$G(\text{DR}) = \Theta \left(\frac{\sqrt{1 - \rho} \Phi^{-1}(\text{DR}) - \Psi^{-1}(\text{PD})}{\sqrt{\rho}} \right)$$

⁸This approach is applied to evaluating the risk of tranches created from mortgages in J. Hull and A. White, “The Risk of Tranches Created from Mortgages,” *Financial Analysts Journal* 66, no. 5 (September/October 2010): 54–67. It provides a better fit to historical data in many situations. Its main disadvantage is that the distributions used are not as easy to deal with as the normal distribution, and numerical analysis may be necessary to determine Ψ and $g(\text{DR})$.

Summary

Risk managers use correlations or covariances to describe the relationship between two variables. The daily covariance rate is the correlation between the daily returns on the variables multiplied by the product of their daily volatilities. The methods for monitoring a covariance rate are similar to those described in Chapter 8 for monitoring a variance rate. Risk managers often try to keep track of a variance–covariance matrix for all the variables to which they are exposed.

The marginal distribution of a variable is the unconditional distribution of the variable. Very often an analyst is in a situation where he or she has estimated the marginal distributions of a set of variables and wants to make an assumption about their correlation structure. If the marginal distributions of the variables happen to be normal, it is natural to assume that the variables have a multivariate normal distribution. In other situations, copulas are used. The marginal distributions are transformed on a percentile-to-percentile basis to normal distributions (or to some other distribution for which there is a multivariate counterpart). The correlation structure between the variables of interest is then defined indirectly from an assumed correlation structure between the transformed variables.

When there are many variables, analysts often use a factor model. This is a way of reducing the number of correlation estimates that have to be made. The correlation between any two variables is assumed to derive solely from their correlations with the factors. The default correlation between different companies can be modeled using a factor-based Gaussian copula model of their times to default.

An important application of copulas for risk managers is to the calculation of the distribution of default rates for loan portfolios. Analysts often assume that a one-factor copula model relates the probability distributions of the times to default for different loans. The percentiles of the distribution of the number of defaults on a large portfolio can then be calculated from the percentiles of the probability distribution of the factor. As we shall see in Chapter 25, this is the approach used in determining credit risk capital requirements for banks under Basel II.

Further Reading

- Cherubini, U., E. Luciano, and W. Vecchiato. *Copula Methods in Finance*. Hoboken, NJ: John Wiley & Sons, 2004.
- Demarta, S., and A. J. McNeil. “The t -Copula and Related Copulas.” Working Paper, Department of Mathematics, ETH Zentrum, Zurich, Switzerland, 2005.
- Engle, R. F., and J. Mezrich. “GARCH for Groups.” *Risk* (August 1996): 36–40.
- Vasicek, O. “Probability of Loss on a Loan Portfolio.” Working Paper, KMV, 1987. (Published in *Risk* in December 2002 under the title “Loan Portfolio Value.”)

Practice Questions and Problems (Answers at End of Book)

- 9.1 If you know the correlation between two variables, what extra information do you need to calculate the covariance?
- 9.2 What is the difference between correlation and dependence? Suppose that $y = x^2$ and x is normally distributed with mean zero and standard deviation one. What is the correlation between x and y ?
- 9.3 What is a factor model? Why are factor models useful when defining a correlation structure between large numbers of variables?
- 9.4 What is meant by a positive-semidefinite matrix? What are the implications of a correlation matrix not being positive-semidefinite?
- 9.5 Suppose that the current daily volatilities of asset A and asset B are 1.6% and 2.5%, respectively. The prices of the assets at close of trading yesterday were \$20 and \$40 and the estimate of the coefficient of correlation between the returns on the two assets made at that time was 0.25. The parameter λ used in the EWMA model is 0.95.
 - (a) Calculate the current estimate of the covariance between the assets.
 - (b) On the assumption that the prices of the assets at close of trading today are \$20.50 and \$40.50, update the correlation estimate.
- 9.6 Suppose that the current daily volatilities of asset X and asset Y are 1.0% and 1.2%, respectively. The prices of the assets at close of trading yesterday were \$30 and \$50 and the estimate of the coefficient of correlation between the returns on the two assets made at this time was 0.50. Correlations and volatilities are updated using a GARCH(1,1) model. The estimates of the model's parameters are $\alpha = 0.04$ and $\beta = 0.94$. For the correlation $\omega = 0.000001$ and for the volatilities $\omega = 0.000003$. If the prices of the two assets at close of trading today are \$31 and \$51, how is the correlation estimate updated?
- 9.7 Suppose that in Problem 8.15 the correlation between the S&P 500 index (measured in dollars) and the FTSE 100 index (measured in sterling) is 0.7, the correlation between the S&P 500 index (measured in dollars) and the dollar-sterling exchange rate is 0.3, and the daily volatility of the S&P 500 index is 1.6%. What is the correlation between the S&P 500 index (measured in dollars) and the FTSE 100 index when it is translated to dollars? (*Hint:* For three variables X , Y , and Z , the covariance between $X + Y$ and Z equals the covariance between X and Z plus the covariance between Y and Z .)
- 9.8 Suppose that two variables V_1 and V_2 have uniform distributions where all values between 0 and 1 are equally likely. Use a Gaussian copula to define the correlation structure between V_1 and V_2 with a copula correlation of 0.3. Produce a table similar to Table 9.5 considering values of 0.25, 0.50, and 0.75 for V_1 and V_2 . (A spreadsheet for calculating the cumulative bivariate normal distribution is on the author's website: www-2.rotman.utoronto.ca/~hull/riskman.)

- 9.9 Assume that you have independent random samples z_1 , z_2 , and z_3 from a standard normal distribution and want to convert them to samples ε_1 , ε_2 , and ε_3 from a trivariate normal distribution using the Cholesky decomposition. Derive three formulas expressing ε_1 , ε_2 , and ε_3 in terms of z_1 , z_2 , and z_3 and the three correlations that are needed to define the trivariate normal distribution.
- 9.10 Explain what is meant by tail dependence. How can you vary tail dependence by the choice of copula?
- 9.11 Suppose that the marginal distributions of V_1 and V_2 are standard normal distributions but that a Student's t -copula with four degrees of freedom and a correlation parameter of 0.5 is used to define the correlation between the variables. How would you obtain samples from the joint distribution?
- 9.12 In Table 9.5, what is the probability density function of V_2 conditional on $V_1 < 0.1$? Compare it with the unconditional distribution of V_2 .
- 9.13 What is the median of the distribution of V_2 when V_1 equals 0.2 in the example in Tables 9.3 and 9.4?
- 9.14 Suppose that a bank has made a large number of loans of a certain type. The one-year probability of default on each loan is 1.5% and the recovery rate is 30%. The bank uses a Gaussian copula for time to default. Use Vasicek's model to estimate the default rate that we are 99.5% certain will not be exceeded. Assume a copula correlation of 0.2.
- 9.15 Suppose that the default rate for a portfolio of consumer loans over the past 10 years has been 1%, 9%, 2%, 3%, 5%, 1%, 6%, 7%, 4%, and 1%. What are the maximum likelihood estimates of the parameters in Vasicek's model?

Further Questions

- 9.16 Suppose that the price of Asset X at close of trading yesterday was \$300 and its volatility was estimated as 1.3% per day. The price of X at the close of trading today is \$298. Suppose further that the price of Asset Y at the close of trading yesterday was \$8, its volatility was estimated as 1.5% per day, and its correlation with X was estimated as 0.8. The price of Y at the close of trading today is unchanged at \$8. Update the volatility of X and Y and the correlation between X and Y using
- The EWMA model with $\lambda = 0.94$
 - The GARCH(1,1) model with $\omega = 0.000002$, $\alpha = 0.04$, and $\beta = 0.94$.
- In practice, is the ω parameter likely to be the same for X and Y?
- 9.17 The probability density function for an exponential distribution is $\lambda e^{-\lambda x}$ where x is the value of the variable and λ is a parameter. The cumulative probability distribution is $1 - e^{-\lambda x}$. Suppose that two variables V_1 and V_2 have exponential distributions with λ parameters of 1.0 and 2.0, respectively. Use a Gaussian copula to define the correlation structure between V_1 and V_2 with a copula correlation of -0.2 . Produce a table similar to Table 9.5 using values of 0.25, 0.5, 0.75, 1, 1.25, and 1.5 for V_1 and V_2 . (A spreadsheet for calculating the cumulative bivariate

normal distribution is on the author's website: www-2.rotman.utoronto.ca/~hull/riskman.)

- 9.18 Create an Excel spreadsheet to produce a chart similar to Figure 9.5 showing samples from a bivariate Student's t -distribution with four degrees of freedom where the correlation is 0.5. Next suppose that the marginal distributions of V_1 and V_2 are Student's t with four degrees of freedom but that a Gaussian copula with a copula correlation parameter of 0.5 is used to define the correlation between the two variables. Construct a chart showing samples from the joint distribution. Compare the two charts you have produced.
- 9.19 Suppose that a bank has made a large number of loans of a certain type. The one-year probability of default on each loan is 1.2%. The bank uses a Gaussian copula for time to default. It is interested in estimating a 99.97% worst case for the percent of loans that default on the portfolio. Show how this varies with the copula correlation.
- 9.20 The default rates in the past 15 years for a certain category of loans are 2%, 4%, 7%, 12%, 6%, 5%, 8%, 14%, 10%, 2%, 3%, 2%, 6%, 7%, and 9%. Use the maximum likelihood method to calculate the best fit values of the parameters in Vasicek's model. What is the probability distribution of the default rate? What is the 99.9% worst case default rate?

Chapter 10

Valuation and Scenario Analysis

Valuation and scenario analysis are two important activities for financial institutions. Both are concerned with estimating future cash flows, but they have different objectives. In valuation, a financial institution is interested in estimating the present value of future cash flows. It does this by calculating the expected (or average) values of the future cash flows across all alternative outcomes and discounting the expected values back to today. In scenario analysis, a financial institution is interested in exploring the full range of situations that might exist at a particular future time. Usually, it is the adverse outcomes that receive the most attention because risk managers working for the financial institution are interested in answering the question: “How bad can things get?”

Suppose that a company sells one million one-year European call options on a stock. The stock price is \$50 and the strike price is \$55. The company might calculate the theoretical value of the options as +\$4.5 million to the buyer and −\$4.5 million to itself. If it sells the options for, say, \$5 million, it can book \$0.5 million of profit. But a scenario analysis might reveal that there is a 5% chance of the stock price rising to above \$80 in one year. This means that there is a 5% chance that the transaction will cost more than \$20 million, after the initial amount received for the options has been taken into account. This example emphasizes the key difference between valuation and scenario analysis. Valuation focuses on what will happen on average. (In our example, \$4.5 million is the present value of the average payoff on the option.) Scenario analysis

focuses on extreme outcomes. (In our example, \$20 million is a possible net cost of the transaction to the company.)

This chapter discusses the way valuation and scenario analysis should be carried out in practice. It distinguishes between real-world projections, which underlie scenario analysis, and risk-neutral projections, which are used for valuation. It shows that risk-neutral valuation can be used for variables such as asset prices that evolve through time and to deal with situations where an outcome depends on whether a particular discrete event occurs (for example, a company defaulting).

10.1 Volatility and Asset Prices

As a preliminary to our discussion of valuation and scenario analysis, it is useful to produce a few results concerned with the behavior of asset prices. Suppose an asset price is currently S_0 . A common assumption is that it has a constant expected growth rate of μ per year (expressed with continuous compounding), and a constant volatility of σ per year.¹ It can be shown that the probability density of the asset price, S_T , at time T years is then given by²

$$\ln S_T \sim \phi[\ln S_0 + (\mu - \sigma^2/2)T, \sigma^2 T] \quad (10.1)$$

where $\phi(m, v)$ denotes a normal distribution with mean m and variance v and $\ln$ is the natural logarithm function. The variable S_T has what is termed a lognormal distribution because its natural logarithm is normally distributed. The mean of $\ln S_T$ is $\ln S_0 + (\mu - \sigma^2/2)T$ and the standard deviation of $\ln S_T$ is $\sigma\sqrt{T}$.

The probability of S_T being less than some value V is the same as the probability of $\ln S_T$ being less than $\ln V$. From the properties of the normal distribution this is

$$\text{Prob}(S_T < V) = N\left[\frac{\ln V - \ln S_0 - (\mu - \sigma^2/2)T}{\sigma\sqrt{T}}\right] = N(-d_2) \quad (10.2)$$

where

$$d_2 = \frac{\ln(S_0/V) + (\mu - \sigma^2/2)T}{\sigma\sqrt{T}}$$

and N is the cumulative normal distribution function (given by NORMSDIST in Excel). The probability that S_T is greater than V at time T is

$$\text{Prob}(S_T > V) = 1 - N(-d_2) = N(d_2) \quad (10.3)$$

Finally, suppose that we want to find the value, V , of S_T that has a probability q of being exceeded. This means that $\text{Prob}(S_T > V) = q$. From equation (10.3), we require $N(d_2) = q$ so that

$$d_2 = \frac{\ln(S_0/V) + (\mu - \sigma^2/2)T}{\sigma\sqrt{T}} = N^{-1}(q)$$

¹Continuous compounding is explained in Appendix A.

²See, for example, J. Hull, *Options, Futures, and Other Derivatives*, 11th ed. Pearson, 2022.

or

$$V = S_0 \exp \left[\left(\mu - \frac{\sigma^2}{2} \right) T - N^{-1}(q) \sigma \sqrt{T} \right] \quad (10.4)$$

where N^{-1} is the inverse of the cumulative normal distribution function (given by NORMSINV in Excel). Similarly, the value, V , of S_T such that $\text{Prob}(S_T < V) = q$ is given by setting $N(-d_2) = q$ or $d_2 = -N^{-1}(q)$. This leads to

$$V = S_0 \exp \left[\left(\mu - \frac{\sigma^2}{2} \right) T + N^{-1}(q) \sigma \sqrt{T} \right] \quad (10.5)$$

10.2 Risk-Neutral Valuation

The single most important result in the valuation of derivatives is risk-neutral valuation. A *risk-neutral world* can be defined as an imaginary world where investors require no compensation for bearing risks. In this world, the required expected return from a risky investment is the same as that for a safe investment. In both cases, the expected return is the risk-free rate. The world we live in is clearly not a risk-neutral world. Investors do require compensation for bearing risks. (A framework for understanding risk-return trade-offs is presented in Chapter 1.) The risk-neutral valuation result, however, states that we can value any derivative by assuming that the world is risk-neutral. We get the right answer not just for the risk-neutral world but for all other worlds as well.

At first blush, the risk-neutral valuation result seems to make no sense. Investors do not live in a risk-neutral world. They do require higher expected returns when the risks they are bearing increase, and this applies to derivatives as well as to other investments.³ But the key point to note when trying to understand risk-neutral valuation is that when we are valuing a derivative we are calculating its value in terms of the price of an underlying asset. (The value of a stock option, for example, is calculated in terms of the price of the underlying stock.) The price of the underlying asset reflects the risk-return trade-offs of market participants. If market participants decide that they require a higher (or lower) return from an asset because of its risk, the price of the asset goes down (or up). What risk-neutral valuation says is that the formula for translating the price of the underlying asset into the value of the derivative is independent of the risk preferences of investors.

In a risk-neutral world, all future expected cash flows are discounted at the risk-free interest rate. This simplifies valuation considerably. Suppose that we are valuing a call option on a stock and the risk-free interest rate is 3%. The steps in implementing risk-neutral valuation are:

1. Assume that the expected (average) future return on the stock is 3%.
2. Calculate the expected payoff from the call option.
3. Discount the expected payoff at 3% to obtain the option's value.

³As explained in Chapter 1, it is systematic (i.e., non-diversifiable) risk that matters to investors.

A natural question is: “Why work in the risk-neutral world when it is more natural to work in the real world?” The answer is that we could in theory value the call option in the real world, but it would be much more difficult. The steps would be

- 1a. Estimate the expected (average) future return on the stock in the real world.
- 2a. Calculate the expected payoff from the call option in the real world.
- 3a. Discount the expected payoff at an appropriate discount rate to obtain the option’s value.

For step 1a, we might be able to come up with a reasonable estimate of the (real-world) expected future return on the stock by estimating its beta and using the capital asset pricing model results in Chapter 1. Step 3a, however, would be really difficult. The correct discount rate to use for the expected payoff in the real world depends on the option’s beta (not the stock’s beta). It is likely to vary during the option’s life. As the stock price changes, the leverage implicit in the option changes and so the discount rate changes. If a single discount rate is used for the whole of the life of a call option, it is surprisingly high. Similarly, if a single discount rate is used for the whole of the life of a put option, it is surprisingly low, usually negative. How do we know this? We can use risk-neutral valuation to value the options and then work back from the answer to see what the correct discount rate would have to be for real-world payoffs to give that answer. We will illustrate this for the case of a binary option shortly.

Risk-neutral valuation is an almost miraculous result. It means that we do not need to concern ourselves with issues such as the riskiness of a derivative and the return required by the market on either the underlying asset or the derivative. All we need to ask ourselves is: If we lived in a world where investors required an expected return equal to the risk-free rate on all investments, how would we value the derivative? Without risk-neutral valuation, the valuation of derivatives would be far more difficult (and far less precise) than it is.

It should be emphasized that risk-neutral valuation (or the assumption that investors do not care about risks when determining expected returns) is nothing more than an artificial device for valuing derivatives. The valuations that are obtained are correct in all worlds, not just the risk-neutral world. When we move from the risk-neutral world to the real world, two things happen. The expected payoffs from the derivative change, and the discount rate that must be used for the payoffs changes. It happens that these two changes always offset each other exactly.

10.2.1 *Application to Forward Contract*

As a simple example of risk-neutral valuation, consider the valuation of a long forward contract on a non-dividend-paying stock. Suppose that the delivery price is K and the time to maturity is T . The value of the contract at maturity is

$$S_T - K$$

where S_T is the stock price at time T . From the risk-neutral valuation argument, the value of the forward contract at time zero (today) is its expected value at time T in a risk-neutral world, discounted at the risk-free rate of interest. Denoting the value of the forward contract by f , this means that

$$f = e^{-rT} \hat{E}(S_T - K)$$

where $\hat{E}$ denotes expected value in a risk-neutral world and r is the risk-free rate (expressed with continuous compounding and assumed constant). Because K is a constant, this becomes

$$f = e^{-rT} \hat{E}(S_T) - Ke^{-rT} \quad (10.6)$$

The expected growth rate of the stock price is r in a risk-neutral world. As a result,

$$\hat{E}(S_T) = S_0 e^{rT}$$

where S_0 is the stock price today. Substituting this into equation (10.6) gives

$$f = S_0 - Ke^{-rT} \quad (10.7)$$

Similarly, the value of a short forward contract is

$$Ke^{-rT} - S_0 \quad (10.8)$$

These results are consistent with Appendix C.

10.2.2 Application to Binary Options

As a further example of the application of risk-neutral valuation, suppose that the price of a non-dividend-paying stock is \$30 and there is a derivative that pays off \$100 in one year if the stock price is greater than \$40 at that time. (This is known as a binary or digital cash-or-nothing call option.) Suppose that the risk-free rate (continuously compounded) is 3% per annum, the expected growth rate of the stock price in the real world is 10% per annum (also continuously compounded), and the stock price volatility is 30% per annum.

In a risk-neutral world, the expected growth of the stock price is 3% per annum. The risk-neutral probability of the stock price being greater than \$40 in one year is obtained by setting $\mu = 0.03$, $T = 1$, $\sigma = 0.3$, $S_0 = 30$, and $V = 40$ in equation (10.3). It is

$$N \left[\frac{\ln(30/40) + (0.03 - 0.3^2/2) \times 1}{0.3 \times \sqrt{1}} \right] = N(-1.0089) = 0.1565$$

The expected payoff from the derivative in a risk-neutral world is therefore $100 \times 0.1565 = \$15.65$. The value of the derivative is calculated by discounting this for one year at the risk-free rate of 3%. It is $15.65e^{-0.03 \times 1}$ or \$15.19.

The real-world probability of the stock price being greater than \$40 in one year is calculated by setting μ equal to the assumed real-world return on the stock, 10%, in

equation (10.3). It is 0.2190. (As will be explained later in this chapter, we do not have to change the volatility when moving from the risk-neutral world to the real world or vice versa.) The expected payoff in the real world is therefore \$21.90. As mentioned earlier, the problem with using the real-world expected payoff for valuation is that we do not know what discount rate to use. The stock price has risk associated with it that is priced by the market (otherwise the expected return on the stock would not be 7% more than the risk-free rate). The derivative has the effect of “leveraging this risk” so that a relatively high discount rate is required for its expected payoff. Because we know the correct value of the derivative is \$15.19, we can deduce that the correct discount rate to apply to the \$21.90 real-world expected payoff must be 36.6%. (This is because $21.90e^{-0.366 \times 1} = 15.19$.)

10.2.3 The Black–Scholes–Merton Application

Consider next a European call option on a non-dividend-paying stock with strike price K and maturity T . Suppose that the risk-free rate is r . The payoff at time T is

$$\max(S_T - K, 0)$$

where S_T is the stock price at time T . The expected payoff in a risk-neutral world is therefore

$$\hat{E}[\max(S_T - K, 0)]$$

where as before $\hat{E}$ denotes expected value in a risk-neutral world. Using risk-neutral valuation, the value of the option is

$$e^{-rT} \hat{E}[\max(S_T - K, 0)] \quad (10.9)$$

Similarly, the value of a put option is

$$e^{-rT} \hat{E}[\max(K - S_T, 0)]$$

After some algebraic manipulations, it can be shown that these equations lead to the Black–Scholes–Merton formulas for European stock options given in Appendix E at the end of the book.⁴

10.2.4 Discrete Outcomes

Risk-neutral valuation can be used when outcomes are discrete. Suppose that one of two mutually exclusive outcomes will occur at time T . Define π_1 as the value of a derivative that pays off \$1 at time T if the first outcome occurs and nothing otherwise. Similarly,

⁴As shown in J. C. Hull, *Options, Futures, and Other Derivatives*, 11th ed. (Pearson, 2022), there are three ways of obtaining the Black–Scholes–Merton formula. One is by deriving the differential equation that must be satisfied by all derivatives and solving it subject to appropriate boundary conditions. Another is to construct a binomial tree for the behavior of the stock price and take the limit as the length of the time step tends to zero. The third approach is to work from equation (10.9). The math for this last approach is in the appendix to Chapter 15 of *Options, Futures, and Other Derivatives*, 11th ed.

define π_2 as the value of a derivative that pays off \$1 at time T if the second outcome occurs and nothing otherwise. By buying both derivatives at a cost of $\pi_1 + \pi_2$, we can be certain to receive \$1 at time T . The value of \$1 received with complete certainty at time T is e^{-RT} where R is the (continuously compounded) risk-free interest rate for maturity T . (We do not have to assume constant interest rates.) It follows that

$$\pi_1 + \pi_2 = e^{-RT} \quad (10.10)$$

Now consider a derivative that at time T provides a payoff of V_1 if the first outcome is realized and V_2 if the second outcome is realized. The value of the derivative is

$$\pi_1 V_1 + \pi_2 V_2$$

This is

$$(\pi_1 + \pi_2) \left(\frac{\pi_1}{\pi_1 + \pi_2} V_1 + \frac{\pi_2}{\pi_1 + \pi_2} V_2 \right)$$

Substituting from equation (10.10), the value of the derivative is

$$e^{-RT} (p_1 V_1 + p_2 V_2)$$

where

$$p_1 = \frac{\pi_1}{\pi_1 + \pi_2} \quad p_2 = \frac{\pi_2}{\pi_1 + \pi_2}$$

From this result it is natural to think of p_1 and p_2 as the risk-neutral probabilities of the two outcomes occurring. The value of a derivative is then the expected payoff in a risk-neutral world discounted at the risk-free rate. This illustrates that risk-neutral valuation applies to discrete outcomes.

The result can be extended to the situation where there are many outcomes. Suppose that one of n mutually exclusive outcomes will occur at time T . Define π_i as the value of a derivative that pays off \$1 if the i th outcome occurs and nothing otherwise ($1 \leq i \leq n$). The value of a derivative that pays V_i for outcome i ($1 \leq i \leq n$) is

$$e^{-RT} \sum_{i=1}^n p_i V_i$$

where p_i , the risk-neutral probability of the i th outcome, is given by

$$p_i = \frac{\pi_i}{\sum_{j=1}^n \pi_j}$$

10.2.5 Application to Default Probabilities

Consider an instrument whose payoff depends on whether a particular company has defaulted. (This type of instrument is known as a *credit derivative*.) The type of analysis we have just presented can be used to show that the derivative should be valued by

1. Estimating risk-neutral default probabilities
2. Calculating the expected payoff from the instrument
3. Discounting the expected payoff at the risk-free rate

As will be explained in Chapter 17, the risk-neutral default probabilities can be implied from the yields on bonds issued by the company or credit default swap spreads. They are in general higher than real-world default probabilities.

10.3 Scenario Analysis

We now move on to consider scenario analysis. Here we are interested in examining what might happen in the future. The objective is not valuation, and future cash flows are not discounted back to today. The world we consider when carrying out a scenario analysis should be the real world, not the risk-neutral world. The risk-neutral world, it should be remembered, is nothing more than an artificial device for valuing derivatives. Risk managers are not normally interested in future outcomes in a hypothetical world where everyone is risk neutral.

Moving between the real world and the risk-neutral world is simplified by a result known as *Girsanov's theorem*. This states that when we move from a world with one set of risk preferences to a world with another set of risk preferences, the expected growth rates of market variables such as stock prices, commodity prices, exchange rates, and interest rates change but their volatilities remain the same.

To illustrate how scenario analysis is carried out, suppose that the expected return on a stock in the real world is 8%. The stock price is currently \$30 per share and its volatility is 25%. Assume you own 10,000 shares. How much could you lose during the next year?

From equation (10.5), the five percentile point of the stock price distribution in one year in the real world is:

$$\$30 \times \exp[(0.08 - 0.25^2/2) \times 1 + N^{-1}(0.05) \times 0.25 \times \sqrt{1}] = \$20.88$$

Similarly, the one percentile point of the stock price distribution is \$17.61.

These results show that you can be 95% certain that you will not lose more than $\$10,000 \times (\$30 - \$20.88)$ or \$91,200 during the next year. Similarly, you can be 99% certain you will not lose more than \$123,900. As we will see in later chapters, these are what are known as *value at risk* estimates.

The key point here is that results are based on the real-world return, not the riskless return. This is because we are conducting a scenario analysis, not a valuation.

10.4 When Both Worlds Have to Be Used

Sometimes a scenario analysis requires us to use both the real world and the risk-neutral world. The real world is used to generate scenarios out to the time horizon being considered. The risk-neutral world is then used to value all outstanding derivative transactions at that time. To take a simple example, suppose a portfolio consists of a single two-year forward contract to sell one million shares of a stock. The stock price is \$50 per share,

the delivery price in the forward contract is \$55, the risk-free rate is 3%, the expected return on the stock is 10%, and its volatility is 30%. We wish to carry out a scenario analysis to investigate what the value of the portfolio might be after six months. In order to do this, we must follow two steps:

1. Calculate a probability distribution for the stock price at the end of six months in the real world.
2. Value the forward contract at the end of six months for the different stock prices that might arise to determine a probability distribution for the contract value at the end of six months. This involves a risk-neutral valuation calculation. (At the end of six months, the contract has 1.5 years remaining.)

Suppose we are interested in a “worst case” outcome in six months where the loss has a probability of only 1% of being exceeded. In this case, because the portfolio is so simple, the worst outcome corresponds to the stock price that has a probability of only 1% of being exceeded in the real world in six months. From equation (10.4) this stock price is

$$50 \exp[(0.1 - 0.3^2/2) \times 0.5 - N^{-1}(0.01) \times 0.3 \times \sqrt{0.5}] = 84.18$$

For this stock price, equation (10.8) gives the value of the forward contract as

$$55e^{-1.5 \times 0.03} - 84.18 = -31.61$$

There is therefore a 1% chance that the portfolio will be worth less than $-\$31.61$ million in six months.

Note that the worst-case stock price was calculated in the real world. The forward contract was then valued for this worst-case outcome using risk-neutral valuation.

10.5 The Calculations in Practice

The example we have just considered is very simple because the portfolio consisted of a single instrument—a two-year short forward contract. We know that the value of this forward contract decreases as the price of the underlying stock increases. When the stock price has only a 1% chance of being exceeded, we know that the forward contract has a value that has only a 1% chance of being worsened.

In practice, a financial institution usually has many instruments in its portfolio and the calculations necessary for a scenario analysis can be quite complicated. It is necessary to generate many scenarios for what might happen in the real world between today and the horizon date and then value the portfolio for each of these scenarios. The loss that has, say, a 1% probability of being exceeded can then be calculated. For example, if 1,000 scenarios are considered, this loss is the 10th worst one.

One approach is to make an assumption about the behavior of the underlying market variables and their correlations. The easiest assumption for the behavior of a market variable over a short time period is from the one in equation (10.1). This gives:

$$\ln(S_{t+\Delta t}) = \ln(S_t) + (\mu - \sigma^2/2)\Delta t + \epsilon\sigma\sqrt{\Delta t}$$

where ϵ is a random sample from a normal distribution with mean zero and standard deviation one. It means that

$$S_{t+\Delta t} = S_t \exp[(\mu - \sigma^2/2)\Delta t + \epsilon\sigma\sqrt{\Delta t}]$$

This equation allows the market variable to be simulated in steps of Δt by sampling from a standard normal distribution.

In the case of short-term interest rates, volatilities, and commodity prices, a more complicated model where the variable exhibits volatility, but is pulled toward a long-run average level is usually assumed. This phenomenon is known as *mean reversion*.

Different market variables are not usually assumed to move independently of each other. Correlations between market variables must be estimated from historical data. These correlations are then reflected in the correlations between the ϵ samples from standard normal distributions. (The way in which samples from a multivariate normal distribution are generated with particular correlations was described in Chapter 9.)

The main problem in carrying out this analysis is determining good estimates for the real-world expected returns of the market variables. In theory, expected returns can be calculated from historical data. In practice, the amount of historical data required to get a reasonably accurate estimate is huge (much greater than that required to get a reasonable estimate of a volatility or a correlation).

One approach to determining the real-world expected return on a stock is to use the capital asset pricing model (see Chapter 1). We first estimate ρ , the correlation of the return on the stock with the return on an index that is representative of the whole market, such as the S&P 500. As explained in Section 1.3, the stock's beta, β , can be estimated as

$$\beta = \rho \frac{\sigma}{\sigma_M}$$

where σ is the volatility of the stock's return and σ_M is the volatility of the S&P 500. The capital asset pricing model can be used to get the return in the real world as

$$R_F + \beta E$$

where R_F is the risk-free rate and E is the expected excess return of the market over the risk-free rate (often assumed to be 5% or 6%).

A similar idea can be used for other variables. Suppose that the volatility of a market variable is σ (which as mentioned is the same in both the real world and the risk-neutral world). The excess of percentage changes in the variable in the real world over those in

the risk-neutral world is $\lambda\sigma$, where λ is a parameter known as the variable's *market price of risk*. In general,

$$\lambda = \frac{\rho}{\sigma_M} E$$

where ρ is the correlation between percentage changes in the value of the variable and returns on the S&P 500.

Consider a commodity price. If its return is uncorrelated with the return on the S&P 500, its expected return can be assumed to be the risk-free rate. Alternatively, if $\rho = 0.3$, $\sigma_M = 0.2$, and $E = 0.06$, we can deduce that $\lambda = 0.09$. If the commodity's price volatility is 40%, its return should be 0.09×0.40 or 3.6% higher in the real world than in the risk-neutral world.

In the case of interest rates, the market price of risk is negative (typically, in the -0.1 to -0.2 range).⁵ This means that in the risk-neutral world interest rates grow faster than they do in the real world. (This makes interest rates different from stock prices, where the reverse is true.)

As we will describe in Chapter 12, a common approach to avoiding the estimation problems we have just mentioned is to use historical simulation. This can be used for short time horizons. It involves assuming that the alternative ways all market variables can behave over the next short period of time are determined by the ways they actually behaved over similar periods of time in the past.

Summary

A confusing aspect of risk management is that valuation and scenario analysis are (or should be) based on different assumptions about how market variables such as stock prices, commodity prices, and exchange rates behave. To value a derivative in terms of the price of the underlying asset, the somewhat artificial assumption that the world is risk neutral is made. This means that the expected return from all assets that are held for investment purposes is assumed to be the risk-free rate and that expected payoffs are discounted at the risk-free rate. The ubiquitous risk-neutral valuation result states that the valuation we obtain when we do this is correct in the real world as well as in the risk-neutral world.

In scenario analysis, we are interested in how market variables behave in the real world (i.e., the world we actually live in). Fortunately, there is a result, Girsanov's theorem, that tells us that the volatility of a variable is the same in the real and risk-neutral worlds. The expected return, however, is liable to be quite different in the two worlds. For example, the expected return from a stock or stock index is quite a bit higher in the real world than in the risk-neutral world. This is because, as discussed in Chapter 1, investors require compensation for bearing risks.

⁵See J. Hull, A. Sokol, and A. White, "Short-Rate Joint-Measure Models," *Risk* (October 2014): 59–63, for an approach to estimating the market price of interest rate risk.

A further confusing point is that sometimes it is necessary to consider both the real world and the risk-neutral world. Consider a financial institution that has a portfolio of derivatives and is interested in how much it could lose over the next year. The financial institution should consider how the relevant market variables will behave in the real world to generate many alternative scenarios for their values in one year. It should then use risk-neutral valuation to determine the value of the portfolio at the one-year point for each of the alternative scenarios.

Further Reading

- Baxter, M., and A. Rennie. *Financial Calculus*. Cambridge: Cambridge University Press, 1996.
- Hull, J. *Options, Futures, and Other Derivatives*, 11th ed. Pearson, 2022.
- Hull, J., A. Sokol, and A. White. “Short-Rate Joint-Measure Models.” *Risk* (October 2014): 59–63.
- Hull, J., and A. White. “Interest Rate Trees: Extensions and Applications.” Working Paper, University of Toronto, 2017.
- Ross, S. “The Recovery Theorem.” *Journal of Finance* 70, no. 2 (April 2015): 615–648.

Practice Questions and Problems (Answers at End of Book)

- 10.1 A stock price has an expected return of 12% and a volatility of 20%. It is currently \$50. What is the probability that it will be greater than \$70 in two years?
- 10.2 In Problem 10.1, what is the stock price that has a 5% probability of being exceeded in two years?
- 10.3 Explain the principle of risk-neutral valuation.
- 10.4 An analyst calculates the expected future value of a stock index in (a) the real world and (b) the risk-neutral world. Which would you expect to be higher? Why?
- 10.5 The value of a derivative that pays off \$100 after one year if a company has defaulted during the year is \$3. The value of a derivative that pays off \$100 after one year if a company has not defaulted is \$95. What is the risk-free rate? What is the risk neutral probability of default?
- 10.6 A binary option pays off \$100 if a stock price is greater than \$30 in three months. The current stock price is \$25 and its volatility is 30%. The risk-free rate is 3% and the expected return on the stock is 10%. What is the value of the option? What is the real-world probability that the payoff will be received?
- 10.7 Explain why it is sometimes necessary to work in both the real world and the risk-neutral world when carrying out a scenario analysis to determine a confidence interval for the value of a portfolio in one year.
- 10.8 Explain the meaning of *mean reversion*.
- 10.9 Explain Girsanov’s theorem.

Further Questions

- 10.10 A stock price has an expected return of 9% and a volatility of 25%. It is currently \$40. What is the probability that it will be less than \$30 in 18 months?
- 10.11 An investor owns 10,000 shares of a particular stock. The current market price is \$80. What is the worst-case value of the portfolio in six months? For the purposes of this question, define the worst-case value of the portfolio as the value that is such that there is only a 1% chance of the actual value being lower. Assume that the expected return on the stock is 8% and its volatility is 20%.
- 10.12 A binary option pays off \$500 if a stock price is greater than \$60 in three months. The current stock price is \$61 and its volatility is 20%. The risk-free rate is 2% and the expected return on the stock is 8%. What is the value of the option? What is the real-world expected payoff?

Part Three

MARKET RISK

Chapter 11

Value at Risk and Expected Shortfall

Value at risk (VaR) and expected shortfall (ES) are attempts to provide a single number that summarizes the total risk in a portfolio. VaR was pioneered by JPMorgan (see Business Snapshot 11.1) and is now widely used by corporate treasurers and fund managers as well as by financial institutions. As we will see in later chapters, it is the measure regulators have traditionally used for many of the calculations they carry out concerned with the setting of capital requirements. But as Chapter 27 will explain, regulators have now switched to ES for some market risk capital calculations.

This chapter introduces the VaR and ES measures and discusses their strengths and weaknesses. Chapters 12 and 13 discuss how they are calculated for market risk while Chapter 19 considers the calculation of VaR for credit risk.

11.1 Definition of VaR

When using the value at risk measure, we are interested in making a statement of the following form:

“We are X percent certain that we will not lose more than V dollars in time T .”

BUSINESS SNAPSHOT 11.1

Historical Perspectives on VaR

JPMorgan is credited with helping to make VaR a widely used measure. The chairman, Dennis Weatherstone, was dissatisfied with the long risk reports he received every day. These contained a huge amount of detail on different measures of the bank's exposures but very little that was really useful to top management. He asked for something simpler that focused on the bank's total exposure over the next 24 hours measured across the bank's entire trading portfolio. At first his subordinates said this was impossible, but eventually they adapted the Markowitz portfolio theory (see Section 1.1) to develop a VaR report. This became known as the 4:15 report because it was placed on the chairman's desk at 4:15 P.M. every day after the close of trading.

Producing the report entailed an enormous amount of work involving the collection of data daily on the positions held by the bank around the world, the handling of different time zones, the estimation of correlations and volatilities, and the development of computer systems. The work was completed in about 1990. The main benefit of the new system was that senior management had a better understanding of the risks being taken by the bank and were better able to allocate capital within the bank. Other banks had been working on similar approaches for aggregating risks and by 1993 VaR was established as an important risk measure.

Banks usually keep the details about the models they develop internally a secret. However, in 1994 JPMorgan made a simplified version of its own system, which they called RiskMetrics, available on the Internet. RiskMetrics included variances and covariances for a very large number of different market variables. This attracted a lot of attention and led to debates about the pros and cons of different VaR models. Software firms started offering their own VaR models, some of which used the RiskMetrics database. After that, VaR was rapidly adopted as a standard by financial institutions and some nonfinancial corporations. The BIS Amendment, which was based on VaR (see Section 25.6), was announced in 1996 and implemented in 1998. Later the RiskMetrics group within JPMorgan was spun off as a separate company. This company developed CreditMetrics for handling credit risks in 1997 and CorporateMetrics for handling the risks faced by non-financial corporations in 1999.

The variable V is the VaR of the portfolio. It is a function of two parameters: the time horizon, T , and the confidence level, X percent. It is the loss level during a time period of length T that we are $X\%$ certain will not be exceeded.

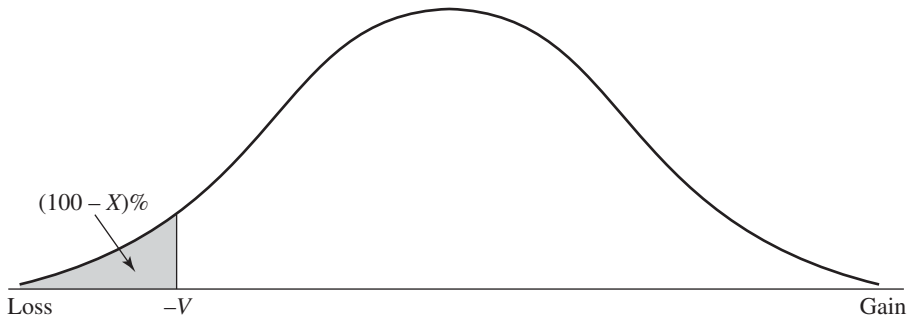


Figure 11.1 Calculation of VaR from the Probability Distribution of the Gain in the Portfolio Value

Losses are negative gains; confidence level is $X\%$; VaR level is V .

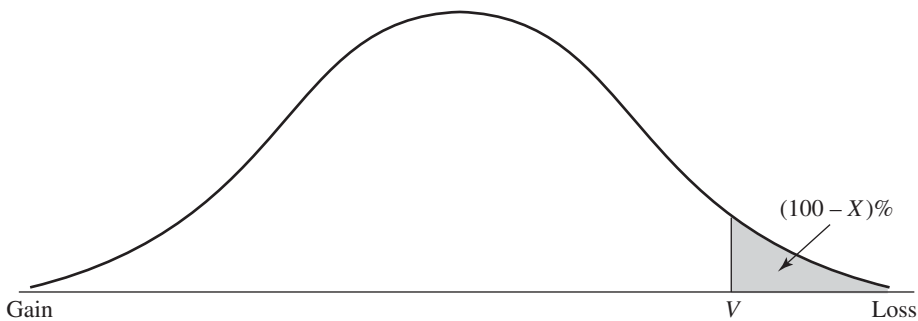


Figure 11.2 Calculation of VaR from the Probability Distribution of the Loss in the Portfolio Value
Gains are negative losses; confidence level is $X\%$; VaR level is V .

VaR can be calculated from either the probability distribution of gains during time T or the probability distribution of losses during time T . (In the former case, losses are negative gains; in the latter case, gains are negative losses.) For example, when T is five days and $X = 97$, VaR is the loss at the 3rd percentile of the distribution of gains over the next five days. Alternatively, it is the loss at the 97th percentile of the distribution of losses over the next five days. More generally, when the distribution of gains is used, VaR is equal to minus the gain at the $(100 - X)$ th percentile of the distribution, as illustrated in Figure 11.1. When the distribution of losses is used, VaR is equal to the loss at the X th percentile of the distribution, as indicated in Figure 11.2.

11.2 Examples of the Calculation of VaR

This section provides four simple examples to illustrate the calculation of VaR. In the first two examples, the probability distribution of the gain (or loss) is a continuous distribution. In the last two examples, it is a discrete distribution.

Example 11.1

Suppose that the gain from a portfolio during six months is normally distributed with a mean of \$2 million and a standard deviation of \$10 million. From the properties of the normal distribution, the one-percentile point of this distribution is $2 - 2.326 \times 10$ or $-\$21.3$ million. The VaR for the portfolio with a time horizon of six months and confidence level of 99% is therefore \$21.3 million.

Example 11.2

Suppose that for a one-year project all outcomes between a loss of \$50 million and a gain of \$50 million are considered equally likely. In this case, the loss from the project has a uniform distribution extending from $-\$50$ million to $+\$50$ million. There is a 1% chance that there will be a loss greater than \$49 million. The VaR with a one-year time horizon and a 99% confidence level is therefore \$49 million.

Example 11.3

A one-year project has a 98% chance of leading to a gain of \$2 million, a 1.5% chance of leading to a loss of \$4 million, and a 0.5% chance of leading to a loss of \$10 million. The cumulative loss distribution is shown in Figure 11.3. The point on this cumulative distribution that corresponds to a cumulative probability of 99% is \$4 million. It follows that VaR with a confidence level of 99% and a one-year time horizon is \$4 million.

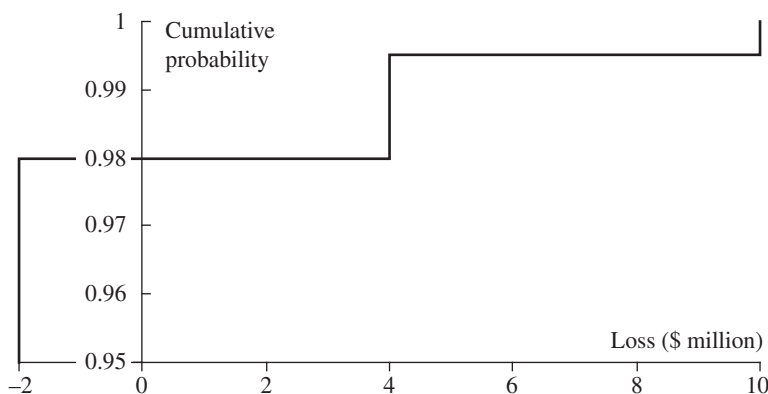


Figure 11.3 Cumulative Loss Distribution for Examples 11.3 and 11.4

Example 11.4

Consider again the situation in Example 11.3. Suppose that we are interested in calculating a VaR using a confidence level of 99.5%. In this case, Figure 11.3 shows that all losses between \$4 million and \$10 million have a probability of 99.5% of not being exceeded. Equivalently, there is a probability of 0.5% of any specified loss level between \$4 million and \$10 million being exceeded. VaR is therefore not uniquely defined. One reasonable convention in this type of situation is to set VaR equal to the midpoint of the range of possible VaR values. This means that, in this case, VaR would equal \$7 million.

11.3 A Drawback of VaR

VaR is an attractive measure because it is easy to understand. In essence, it asks the simple question “How bad can things get?” This is the question all senior managers want answered. They are very comfortable with the idea of compressing risks associated with all the market variables underlying a portfolio into a single number.

However, when VaR is used in an attempt to limit the risks taken by a trader, it can lead to undesirable results. Suppose that a bank tells a trader that the one-day 99% VaR of the trader’s portfolio must be limited to \$10 million. The trader can construct a portfolio where there is a 99.1% chance that the daily loss is less than \$10 million and a 0.9% chance that it is \$500 million. The trader is satisfying the risk limits imposed by the bank but is clearly taking unacceptable risks. The sort of probability distribution of gains that the trader might aim for is shown in Figure 11.4. The VaR in Figure 11.4 might be the same as the VaR in Figure 11.1. But the portfolio in Figure 11.4 is much riskier than the portfolio in Figure 11.1 because a large loss is more likely.

It might be thought that a probability distribution such as that in Figure 11.4 would never occur in practice. In fact it is not that unusual. Many trading strategies give a high

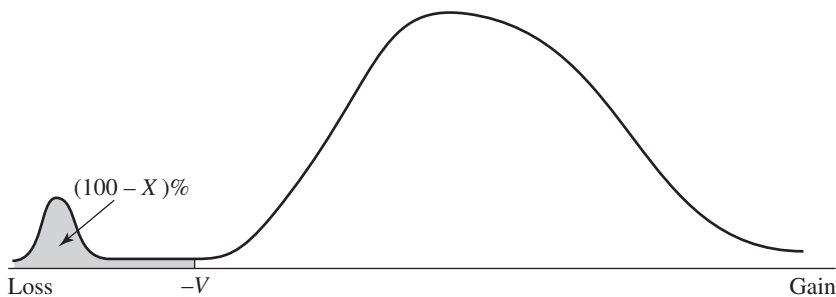


Figure 11.4 Probability Distribution for Gain in Portfolio Value during Time T . Confidence Level is $X\%$. Portfolio has the same VaR level, V , as in Figure 11.1, but a large loss is more likely.

probability of good returns and a small probability of a huge loss. (For example, writing out-of-the-money options is a strategy where most of the time the trader collects the option premium and does not have to provide a payoff to the option buyer. But occasionally the option is exercised in circumstances where the trader takes a big loss.) Many traders like taking high risks in the hope of realizing high returns. If they can find ways of taking high risks without violating risk limits, they will do so. To quote one trader the author has talked to: “I have never met a risk control system that I cannot trade around.”

11.4 Expected Shortfall

A measure that can produce better incentives for traders than VaR is *expected shortfall* (ES). This is also sometimes referred to as *conditional value at risk*, *conditional tail expectation*, or *expected tail loss*. Whereas VaR asks the question: “How bad can things get?” ES asks: “If things do get bad, what is the expected loss?” ES, like VaR, is a function of two parameters: T (the time horizon) and X (the confidence level). Indeed, in order to calculate ES it is necessary to calculate VaR first. ES is the expected loss during time T conditional on the loss being greater than the VaR. For example, suppose that $X = 99$, T is 10 days, and the VaR is \$64 million. The ES is the average amount lost over a 10-day period assuming that the loss is greater than \$64 million.

Setting an ES limit rather than a VaR limit for traders makes it less likely that they will be able to take the sort of position indicated by Figure 11.4. Also, as shown in the next section, ES has better properties than VaR in that it always recognizes the benefits of diversification. One disadvantage is that it does not have the simplicity of VaR and as a result is more difficult to understand. Another is that it is more difficult to back-test a procedure for calculating ES than it is to back-test a procedure for calculating VaR. (Back-testing, as will be explained later, is a way of looking at historical data to test the reliability of a particular methodology for calculating a risk measure.)

11.5 Coherent Risk Measures

Suppose that the VaR of a portfolio for a confidence level of 99.9% and a time horizon of one year is \$50 million. This means that in extreme circumstances (theoretically, once every thousand years) the financial institution will lose more than \$50 million in a year. It also means that if it keeps \$50 million in capital it will have a 99.9% probability of not running out of capital in the course of one year.

Suppose we are trying to design a risk measure that will equal the capital a financial institution is required to keep. Is VaR (with an appropriate time horizon and an appropriate confidence level) the best measure? Artzner et al. have examined this question. They

first proposed a number of properties that such a risk measure should have.¹ These are:

1. *Monotonicity*: If a portfolio produces a worse result than another portfolio for every state of the world, its risk measure should be greater.
2. *Translation Invariance*: If an amount of cash K is added to a portfolio, its risk measure should go down by K .
3. *Homogeneity*: Changing the size of a portfolio by a factor λ while keeping the relative amounts of different items in the portfolio the same should result in the risk measure being multiplied by λ .
4. *Subadditivity*: The risk measure for two portfolios after they have been merged should be no greater than the sum of their risk measures before they were merged.

The first condition is straightforward. If one portfolio always performs worse than another portfolio, it clearly should be viewed as more risky and require more capital. The second condition is also reasonable. If we add an amount of cash equal to K to a portfolio, the cash provides a buffer against losses and should reduce the capital requirement by K . The third condition is also reasonable. If we double the size of a portfolio, presumably we should require twice as much capital.² The fourth condition states that diversification helps reduce risks. When we aggregate two portfolios, the total risk measure should either decrease or stay the same.

VaR satisfies the first three conditions. However, it does not always satisfy the fourth one, as is illustrated by the following two examples.

Example 11.5

Suppose each of two independent projects has a probability of 0.02 of a loss of \$10 million and a probability of 0.98 of a loss of \$1 million during a one-year period. The one-year, 97.5% VaR for each project is \$1 million. When the projects are put in the same portfolio, there is a $0.02 \times 0.02 = 0.0004$ probability of a loss of \$20 million, a $2 \times 0.02 \times 0.98 = 0.0392$ probability of a loss of \$11 million, and a $0.98 \times 0.98 = 0.9604$ probability of a loss of \$2 million. The one-year 97.5% VaR for the portfolio is \$11 million. The total of the VaRs of the projects considered separately is \$2 million. The VaR of the portfolio is therefore greater than the sum of the VaRs of the projects by \$9 million. This violates the subadditivity condition.

¹See P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Coherent Measures of Risk," *Mathematical Finance* 9 (1999): 203–228.

²This is true provided the portfolio is not too large. As a portfolio's size increases, it becomes less liquid and proportionally more capital may be required.

Example 11.6

A bank has two \$10 million one-year loans. The probabilities of default are as indicated in the following table.

Outcome	Probability
Neither loan defaults	97.50%
Loan 1 defaults; Loan 2 does not default	1.25%
Loan 2 defaults; Loan 1 does not default	1.25%
Both loans default	0.00%

If a default occurs, all losses between 0% and 100% of the principal are equally likely. If the loan does not default, a profit of \$0.2 million is made.

Consider first Loan 1. This has a 1.25% chance of defaulting. When a default occurs the loss experienced is evenly distributed between zero and \$10 million. This means that there is a 1.25% chance that a loss greater than zero will be incurred; there is a 0.625% chance that a loss greater than \$5 million is incurred; there is no chance of a loss greater than \$10 million. The loss level that has a probability of 1% of being exceeded is \$2 million. (Conditional on a loss being made, there is an 80% or 0.8 chance that the loss will be greater than \$2 million. Because the probability of a loss is 1.25% or 0.0125, the unconditional probability of a loss greater than \$2 million is $0.8 \times 0.0125 = 0.01$ or 1%.) The one-year 99% VaR is therefore \$2 million. The same applies to Loan 2.

Consider next a portfolio of the two loans. There is a 2.5% probability that a default will occur. As before, the loss experienced on a defaulting loan is evenly distributed between zero and \$10 million. The VaR in this case turns out to be \$5.8 million. This is because there is a 2.5% (0.025) chance of one of the loans defaulting and conditional on this event is a 40% (0.4) chance that the loss on the loan that defaults is greater than \$6 million. The unconditional probability of a loss from a default being greater than \$6 million is therefore $0.4 \times 0.025 = 0.01$ or 1%. In the event that one loan defaults, a profit of \$0.2 million is made on the other loan, showing that the one-year 99% VaR is \$5.8 million.

The total VaR of the loans considered separately is $2 + 2 = \$4$ million. The total VaR after they have been combined in the portfolio is \$1.8 million greater at \$5.8 million. This shows that the subadditivity condition is violated. (This is in spite of the fact that there are clearly very attractive diversification benefits from combining the loans into a single portfolio—particularly because they cannot default together.)

Risk measures satisfying all four conditions given above are referred to as coherent. Examples 11.5 and 11.6 illustrate that VaR is not coherent. It can be shown that the ES measure is always coherent. The following examples illustrate this.

Example 11.7

Consider again the situation in Example 11.5. The VaR for one of the projects considered on its own is \$1 million. To calculate the ES for a 97.5% confidence level we note that, of the 2.5% tail of the loss distribution, 2% corresponds to a \$10 million loss and 0.5% to a \$1 million loss. (Note that the other 97.5% of the distribution also corresponds to a loss of \$1 million.) Conditional that we are in the 2.5% tail of the loss distribution, there is therefore an 80% probability of a loss of \$10 million and a 20% probability of a loss of \$1 million. The expected loss is $0.8 \times 10 + 0.2 \times 1$ or \$8.2 million.

When the two projects are combined, of the 2.5% tail of the loss distribution, 0.04% corresponds to a loss of \$20 million and 2.46% corresponds to a loss of \$11 million. Conditional that we are in the 2.5% tail of the loss distribution, the expected loss is therefore $(0.04/2.5) \times 20 + (2.46/2.5) \times 11$ or \$11.144 million. This is the ES.

Because $8.2 + 8.2 > 11.144$, the ES measure does satisfy the subadditivity condition for this example.

Example 11.8

Consider again the situation in Example 11.6. We showed that the VaR for a single loan is \$2 million. The ES from a single loan when the time horizon is one year and the confidence level is 99% is therefore the expected loss on the loan conditional on a loss greater than \$2 million. Given that losses are uniformly distributed between zero and \$10 million, the expected loss conditional on a loss greater than \$2 million is halfway between \$2 million and \$10 million, or \$6 million.

The VaR for a portfolio consisting of the two loans was calculated in Example 11.6 as \$5.8 million. The ES from the portfolio is therefore the expected loss on the portfolio conditional on the loss being greater than \$5.8 million. When one loan defaults, the other (by assumption) does not and outcomes are uniformly distributed between a gain of \$0.2 million and a loss of \$9.8 million. The expected loss, given that we are in the part of the distribution between \$5.8 million and \$9.8 million, is \$7.8 million. This is therefore the ES of the portfolio.

Because \$7.8 million is less than $2 \times \$6$ million, the ES measure does satisfy the subadditivity condition.

The subadditivity condition is not of purely theoretical interest. Occasionally a financial institution finds that, when it combines two portfolios (e.g., its equity portfolio and its fixed-income portfolio), the total VaR goes up.

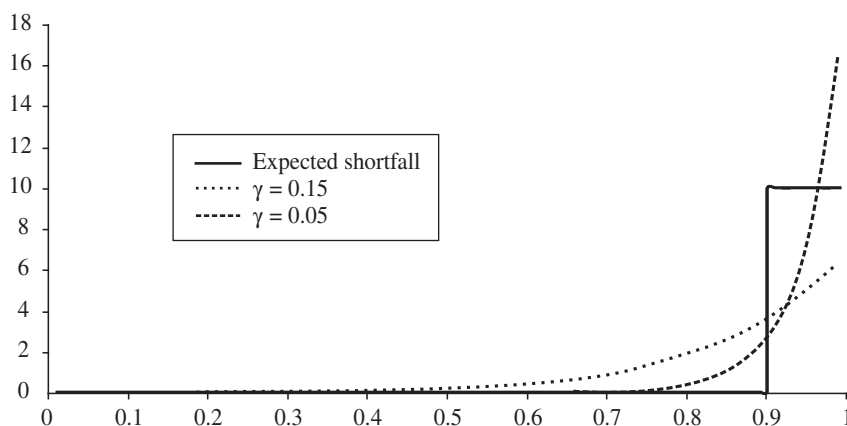


Figure 11.5 Weights as a Function of Percentiles
 (a) Expected shortfall when $X = 90\%$, (b) exponential spectral risk measure with $\gamma = 0.15$, and
 (c) exponential spectral risk measure with $\gamma = 0.05$.

11.5.1 Spectral Risk Measures

A risk measure can be characterized by the weights it assigns to percentiles of the loss distribution.³ VaR gives a 100% weighting to the X th percentiles and zero to other percentiles. ES gives equal weight to all percentiles greater than the X th percentile and zero weight to all percentiles below the X th percentile. We can define other risk measures by making other assumptions about the weights assigned to percentiles. A general result is that a risk measure is coherent (i.e., it satisfies the subadditivity condition) if the weight assigned to the q th percentile of the loss distribution is a nondecreasing function of q . ES satisfies this condition. However, VaR does not, because the weights assigned to percentiles greater than X are less than the weight assigned to the X th percentile. Some researchers have proposed measures where the weights assigned to the q th percentile of the loss distribution increase relatively fast with q . One idea is to make the weight assigned to the q th percentile proportional to $e^{-(1-q)/\gamma}$ where γ is a constant. This is referred to as the *exponential spectral risk measure*. Figure 11.5 shows the weights assigned to loss percentiles for ES and for the exponential spectral risk measure when γ has two different values.

11.6 Choice of Parameters for VaR and ES

For VaR and ES, a user must choose two parameters: the time horizon and the confidence level. A simple assumption is that the change in the portfolio value at the time horizon is normally distributed. As explained in Section 8.3, this is not always a good assumption.

³Percentiles are also referred to as quantiles or fractiles.

However, it is useful for us to consider the consequences of the assumption at this stage. When the loss in the portfolio value has a mean of μ and a standard deviation of σ ,

$$\text{VaR} = \mu + \sigma N^{-1}(X) \quad (11.1)$$

where X is the confidence level and $N^{-1}(\cdot)$ is the inverse cumulative normal distribution (which can be calculated using NORMSINV in Excel). For relatively short time horizons, μ is often assumed to be zero. VaR for a particular confidence level is then proportional to σ .

Example 11.9

Suppose that the loss from a portfolio over a 10-day time horizon is normal with a mean of zero and a standard deviation of \$20 million. The 10-day 99% VaR is

$$20N^{-1}(0.99) = 46.5$$

or \$46.5 million.

When the loss is assumed to be normally distributed with mean μ and standard deviation σ , ES with a confidence level of X is given by

$$ES = \mu + \sigma \frac{e^{-Y^2/2}}{\sqrt{2\pi}(1-X)} \quad (11.2)$$

where Y is the X th percentile point of the standard normal distribution (i.e., it is the point on a normal distribution with mean zero and standard deviation one that has a probability $1 - X$ of being exceeded). This shows that when μ is assumed to be zero, ES, like VaR, is proportional to σ .

Example 11.10

Consider again the situation in Example 11.9 where the loss from a portfolio over a 10-day time horizon is normally distributed with a mean of zero and a standard deviation of \$20 million. Because 2.326 is the point on a standard normal distribution that has a 1% chance of being exceeded, the 10-day 99% ES is

$$20 \frac{e^{-2.326^2/2}}{\sqrt{2\pi} \times 0.01} = 53.3$$

or \$53.3 million.

11.6.1 The Time Horizon

An appropriate choice for the time horizon, when VaR or ES is calculated, depends on the application. When positions are very liquid and actively traded, it makes sense to use a short time horizon (perhaps only a few days). If the measure calculated turns out to be unacceptable, the portfolio can be adjusted fairly quickly. Also, a longer time horizon might not be meaningful because of changes in the composition of the portfolio.

When VaR or ES is being calculated by the manager of a pension fund, a longer time horizon is likely to be used. This is because the portfolio is traded less actively and some of the instruments in the portfolio are less liquid. When the liquidity of a portfolio varies from one instrument to another, the definition of VaR or ES can be changed so that the changes considered vary from one market variable to another. Consider, for example, a portfolio consisting of shares in IBM and a corporate bond that trades fewer than 10 times per year. It could make sense to calculate a risk measure from the change in the price of IBM that we are 99% confident will not be exceeded over 10 days and the change in the bond price that we are 99% certain will not be exceeded over 60 days. This is the approach used by regulators in the Fundamental Review of the Trading Book, which is discussed in Chapter 27.

Whatever the application, when market risks are being considered, analysts often start by calculating VaR or ES for a time horizon of one day. The usual assumption is

$$T\text{-day VaR} = 1\text{-day VaR} \times \sqrt{T} \quad (11.3)$$

$$T\text{-day ES} = 1\text{-day ES} \times \sqrt{T} \quad (11.4)$$

These formulas are exactly true when the changes in the value of the portfolio on successive days have independent identical normal distributions with mean zero. In other cases, they are approximations. The formulas follow from equations (11.1) and (11.2) and the following results.

1. The standard deviation of the sum on T independent identical distributions is $\sqrt{T}$ times the standard deviation of each distribution.
2. The sum of the independent normal distributions is normal.

11.6.2 Impact of Autocorrelation

In practice, the changes in the value of a portfolio from one day to the next are not always totally independent. Define ΔP_i as the change in the value of a portfolio on day i . A simple assumption is first-order autocorrelation where the correlation between ΔP_i and ΔP_{i-1} is ρ for all i . Suppose that the variance of ΔP_i is σ^2 for all i . Using the usual formula for the variance of the sum of two variables, the variance of $\Delta P_{i-1} + \Delta P_i$ is

$$\sigma^2 + \sigma^2 + 2\rho\sigma^2 = 2(1 + \rho)\sigma^2$$

Table 11.1 Ratio of T -Day VaR (ES) to One-Day VaR (ES) for Different Values of T When There Is First-Order Correlation and Daily Changes Have Identical Normal Distributions with Mean Zero.

	$T = 1$	$T = 2$	$T = 5$	$T = 10$	$T = 50$	$T = 250$
$\rho = 0$	1.00	1.41	2.24	3.16	7.07	15.81
$\rho = 0.05$	1.00	1.45	2.33	3.31	7.43	16.62
$\rho = 0.1$	1.00	1.48	2.42	3.46	7.80	17.47
$\rho = 0.2$	1.00	1.55	2.62	3.79	8.62	19.35

The correlation between ΔP_{i-j} and ΔP_i is ρ^j . Extending the analysis leads to the following formula for the standard deviation of $\sum_{i=1}^T \Delta P_i$ (see Problem 11.11):

$$\sigma \sqrt{T + 2(T-1)\rho + 2(T-2)\rho^2 + 2(T-3)\rho^3 + \dots + 2\rho^{T-1}} \quad (11.5)$$

Table 11.1 shows the impact of autocorrelation on the ratio of the T -day VaR (ES) to the one-day VaR (ES). It assumes that the distributions of daily changes in the portfolio are identical normals with mean zero. Note that the ratio of the T -day VaR (ES) to the one-day VaR (ES) does not depend on the daily standard deviation, σ , or the confidence level. This follows from the results in equations (11.1) and (11.2) and the property of equation (11.5) that the T -day standard deviation is proportional to the one-day standard deviation. Comparing the $\rho = 0$ row in Table 11.1 with the other rows shows that the existence of autocorrelation results in the VaR and ES estimates calculated from equations (11.3) and (11.4) being too low.

Example 11.11

Suppose that daily changes in a portfolio value are normally distributed with mean zero and standard deviation \$3 million. The first-order autocorrelation of daily changes is 0.1. From equation (11.5), the standard deviation of the change in the portfolio value over five days is

$$3\sqrt{5 + 2 \times 4 \times 0.1 + 2 \times 3 \times 0.1^2 + 2 \times 2 \times 0.1^3 + 2 \times 1 \times 0.1^4} = 7.265$$

The five-day 95% VaR is therefore $7.265 \times N^{-1}(0.95) = 11.95$ or \$11.95 million. The five-day ES is

$$7.265 \times \frac{e^{-1.645^2/2}}{\sqrt{2\pi} \times 0.05} = 14.98$$

Note that the ratio of the five-day standard deviation of portfolio changes to the one-day standard deviation is $7.265/3 = 2.42$. This is the number in Table 11.1 for $\rho = 0.1$ and $T = 5$.

11.6.3 Confidence Level

The confidence level chosen for VaR or ES is likely to depend on a number of factors. Suppose that a bank wants to maintain an AA credit rating and knows from statistics produced by rating agencies that companies with this credit rating have a 0.02% chance of defaulting over a one-year period. It might choose to use a 99.98% confidence level in conjunction with a one-year time horizon when calculating VaR for internal risk management purposes. Suppose, for example, that the one-year 99.98% VaR across all exposures is \$5 billion. This means that with \$5 billion of capital the bank will have a 0.02% chance of becoming insolvent (i.e., running out of equity) during one year. This type of analysis might be communicated by banks to rating agencies in an attempt to convince the rating agency that the bank deserves its AA rating.

The confidence level that is actually used for the first VaR or ES calculation is sometimes much less than the one that is required. This is because it is very difficult to estimate a VaR directly when the confidence level is very high. A general approach for increasing the confidence level is extreme value theory, discussed in the next chapter. If daily portfolio changes are assumed to be normally distributed with zero mean, we can use equations (11.1) and (11.2) to convert a VaR or ES calculated with one confidence level to that with another confidence level. For example, suppose that σ is the standard deviation of the change in the portfolio value over a certain time horizon and that the expected change in the portfolio value is zero. Denote VaR and ES for a confidence level of X by $\text{VaR}(X)$ and $\text{ES}(X)$, respectively. From equation (11.1)

$$\text{VaR}(X) = \sigma N^{-1}(X)$$

for all confidence levels X . It follows that a VaR with a confidence level of X^* can be calculated from a VaR with a lower confidence level of X using

$$\text{VaR}(X^*) = \text{VaR}(X) \frac{N^{-1}(X^*)}{N^{-1}(X)} \quad (11.6)$$

Similarly, from equation (11.2)

$$\text{ES}(X^*) = \text{ES}(X) \frac{(1 - X)e^{-(Y^* - Y)(Y^* + Y)/2}}{1 - X^*} \quad (11.7)$$

where Y and Y^* are the points on the standard normal distribution that have probabilities $1 - X$ and $1 - X^*$ of being exceeded.

Equations (11.6) and (11.7) assume that the two VaR/ES measures have the same time horizon. If we want to change the time horizon and the confidence level, we can use the equations in conjunction with equation (11.3) or (11.4).

Example 11.12

Suppose that the one-day VaR with a confidence level of 95% is \$1.5 million and the one-day expected shortfall is \$2 million. Using the assumption that the distribution of changes in the portfolio value is normal with mean zero, equation (11.6) gives the one-day 99% VaR as

$$1.5 \times \frac{2.326}{1.645} = 2.12$$

or \$2.12 million. Equation (11.7) gives the one-day 99% ES as

$$2 \times \frac{0.05}{0.01} e^{-(2.326-1.645) \times (2.326+1.645)/2} = 2.58$$

or \$2.58 million.

11.7 Marginal, Incremental, and Component Measures

Consider a portfolio that is composed of a number of subportfolios. The subportfolios could correspond to asset classes (e.g., domestic equities, foreign equities, fixed income, and derivatives). They could correspond to the different business units (e.g., retail banking, investment banking, and proprietary trading). They could even correspond to individual trades. Analysts sometimes calculate measures of the contribution of each subportfolio to VaR or ES.

Suppose that the amount invested in i th subportfolio is x_i . The *marginal value at risk* for the i th subportfolio is the sensitivity of VaR to the amount invested in the i th subportfolio. It is

$$\frac{\partial \text{VaR}}{\partial x_i}$$

To estimate marginal VaR, we can increase x_i to $x_i + \Delta x_i$ for a small Δx_i and recalculate VaR. If ΔVaR is the increase in VaR, the estimate of marginal VaR is $\Delta \text{VaR} / \Delta x_i$. For a well-diversified investment portfolio, marginal VaR is closely related to the capital asset pricing model's beta (see Section 1.3). If an asset's beta is high, its marginal VaR will tend to be high. If its beta is low, the marginal VaR tends to be low. In some circumstances, marginal VaR is negative indicating that increasing the weighting of a particular subportfolio reduces the risk of the portfolio.

The *incremental value at risk* for the i th subportfolio is the incremental effect of the i th subportfolio on VaR. It is the difference between VaR with the subportfolio and VaR without the subportfolio. Traders are often interested in the incremental VaR for a new trade.

The *component value at risk* for the i th subportfolio is

$$C_i = \frac{\partial \text{VaR}}{\partial x_i} x_i \quad (11.8)$$

This can be approximated as

$$\frac{\Delta \text{VaR}}{\Delta x_i} x_i$$

It can be calculated by making a small percentage change $y_i = \Delta x_i / x_i$ in the amount invested in the i th subportfolio and recalculating VaR. If ΔVaR is the increase in VaR, the estimate of component VaR is $\Delta \text{VaR} / y_i$. In many situations, component VaR is a reasonable approximation to incremental VaR. This is because, if a subportfolio is small in relation to the size of the whole portfolio, it can be assumed that the marginal VaR remains constant as x_i is reduced all the way to zero. When this assumption is made, the impact of reducing x_i to zero is x_i times the marginal VaR—which is the component VaR.

Marginal ES, incremental ES, and component ES can be defined similarly to marginal VaR, incremental VaR, and component VaR, respectively.

11.8 Euler's Theorem

A result produced by the great mathematician Leonhard Euler many years ago turns out to be very important when a risk measure for a whole portfolio is allocated to subportfolios. Suppose that V is a risk measure for a portfolio and x_i is a measure of the size of the i th subportfolio ($1 \leq i \leq M$). Assume that, when x_i is changed to λx_i for all x_i (so that the size of the portfolio is multiplied by λ), V changes to λV . This corresponds to the third condition in Section 11.5 and is known as linear homogeneity. It is true for most risk measures.⁴

Euler's theorem shows that it is then true that

$$V = \sum_{i=1}^M \frac{\partial V}{\partial x_i} x_i \quad (11.9)$$

This result provides a way of allocating V to the subportfolios.

When the risk measure is VaR, Euler's theorem gives

$$\text{VaR} = \sum_{i=1}^M C_i$$

where C_i , as in equation (11.8), is the component VaR for the i th subportfolio. This shows that the total VaR for a portfolio is the sum of the component VaRs for the subportfolios. Component VaRs are therefore a convenient way of allocating a total VaR to subportfolios. As explained in the previous section, component VaRs also have the attractive property that the i th component VaR for a large portfolio is approximately equal to the incremental VaR for that component.

⁴An exception could be a risk measure that incorporates liquidity. As a portfolio becomes larger, its liquidity declines.

When the risk measure is ES, Euler's theorem similarly shows that the total ES is the sum of the component ESs:

$$ES = \sum_{i=1}^M \frac{\partial ES}{\partial x_i} x_i$$

ES can therefore be allocated to the component parts of a business similarly to VaR. In Chapter 28, we will show how Euler's theorem is used to allocate a bank's economic capital to its business units.

Euler's theorem allows risk to be decomposed into its components. It is a useful tool in determining risk in what is referred to as *risk budgeting*. This is concerned with the amount of risk that should be allocated to different components of a portfolio. If Euler's decomposition shows that an unacceptable percentage of the risk is attributable to a particular component, the portfolio can be rebalanced.

11.9 Aggregating VaRs and ESs

Sometimes a business has calculated VaRs, with the same confidence level and time horizon, for several different segments of its operations and is interested in aggregating them to calculate a total VaR. A formula for doing this is

$$VaR_{\text{total}} = \sqrt{\sum_i \sum_j VaR_i VaR_j \rho_{ij}} \quad (11.10)$$

where VaR_i is the VaR for the i th segment, VaR_{total} is the total VaR, and ρ_{ij} is the correlation between losses from segment i and segment j . This is exactly true when the losses (gains) have zero-mean normal distributions and provides a good approximation in many other situations. The same is true when VaR is replaced by ES in equation (11.10).

Example 11.13

Suppose the ESs calculated for two segments of a business are \$60 million and \$100 million. The correlation between the losses is estimated as 0.4. An estimate of the total ES is

$$\sqrt{60^2 + 100^2 + 2 \times 60 \times 100 \times 0.4} = 135.6$$

11.10 Back-Testing

Back-testing is an important reality check for a risk measure. It is a test of how well the current procedure for calculating the measure would have worked in the past. VaR is easier to back-test than ES. No doubt this is one of the reasons why regulators have in the past been reluctant to switch from VaR to ES for market risk. As we will explain

in Chapter 27 they now use ES to determine regulatory capital while back-testing using VaR estimates.

Suppose that we have developed a procedure for calculating a one-day 99% VaR. Back-testing involves looking at how often the loss in a day would have exceeded the one-day 99% VaR when the latter is calculated using the current procedure. Days when the actual loss exceeds VaR are referred to as *exceptions*. If exceptions happen on about 1% of the days, we can feel reasonably comfortable with the current methodology for calculating VaR. If they happen on, say, 7% of days, the methodology is suspect and it is likely that VaR is underestimated. From a regulatory perspective, the capital calculated using the current VaR estimation procedure is then too low. On the other hand, if exceptions happen on, say, 0.3% of days, it is likely that the current procedure is overestimating VaR and the capital calculated is too high.

One issue in back-testing a one-day VaR is whether we should consider changes made in the portfolio during a day. There are two possibilities. The first is to compare VaR with the hypothetical change in the portfolio value calculated on the assumption that the composition of the portfolio remains unchanged during the day. The other is to compare VaR to the actual change in the value of the portfolio during the day. The assumption underlying the calculation of VaR is that the portfolio will remain unchanged during the day and so the first comparison based on hypothetical changes is more theoretically correct. However, it is actual changes in the portfolio value that we are ultimately interested in. In practice, risk managers usually compare VaR to both hypothetical portfolio changes and actual portfolio changes (and regulators usually insist on seeing the results of back-testing using both types of changes). The actual changes are adjusted for items unrelated to the market risk, such as fee income and profits from trades carried out at prices different from the mid-market.

Suppose that the confidence level for a one-day VaR is $X\%$. If the VaR model used is accurate, the probability of the VaR being exceeded on any given day is $p = 1 - X/100$. Suppose that we look at a total of n days and we observe that the VaR level is exceeded on m of the days where $m/n > p$. Should we reject the model for producing values of VaR that are too low? Expressed formally, we can consider two alternative hypotheses:

1. The probability of an exception on any given day is p .
2. The probability of an exception on any given day is greater than p .

From the properties of the binomial distribution, the probability of the VaR level being exceeded on m or more days is

$$\sum_{k=m}^n \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

This can be calculated using the BINOMDIST function in Excel. An often-used significance level in statistical tests is 5%. If the probability of the VaR level being exceeded on m or more days is less than 5%, we reject the first hypothesis that the probability of

an exception is p . If the probability of the VaR level being exceeded on m or more days is greater than 5%, the hypothesis is not rejected.

Example 11.14

Suppose that we back-test a VaR model using 600 days of data. The VaR confidence level is 99% and we observe nine exceptions. The expected number of exceptions is six. Should we reject the model? The probability of nine or more exceptions can be calculated in Excel as $1 - \text{BINOMDIST}(8, 600, 0.01, \text{TRUE})$. It is 0.152. At a 5% significance level we should not therefore reject the model. However, if the number of exceptions had been 12 we would have calculated the probability of 12 or more exceptions as 0.019 and rejected the model. The model is rejected when the number of exceptions is 11 or more. (The probability of 10 or more exceptions is greater than 5%, but the probability of 11 or more is less than 5%.)

When the number of exceptions, m , is lower than the expected number of exceptions, we can similarly test whether the true probability of an exception is 1%. (In this case, our alternative hypothesis is that the true probability of an exception is less than 1%.) The probability of m or fewer exceptions is

$$\sum_{k=0}^m \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

and this is compared with the 5% threshold.

Example 11.15

Suppose again that we back-test a VaR model using 600 days of data when the VaR confidence level is 99% and we observe one exception, well below the expected number of six. Should we reject the model? The probability of one or zero exceptions can be calculated in Excel as $\text{BINOMDIST}(1, 600, 0.01, \text{TRUE})$. It is 0.017. At a 5% significance level, we should therefore reject the model. However, if the number of exceptions had been two or more, we would not have rejected the model.

The tests we have considered so far have been one-tailed tests. In Example 11.14, we assumed that the true probability of an exception was either 1% or greater than 1%. In Example 11.15, we assumed that it was 1% or less than 1%. Kupiec (1995) has proposed a relatively powerful two-tailed test.⁵ If the probability of an exception under the VaR model is p and m exceptions are observed in n trials, then

$$-2 \ln[(1-p)^{n-m} p^m] + 2 \ln[(1-m/n)^{n-m} (m/n)^m] \quad (11.11)$$

⁵See P. Kupiec, "Techniques for Verifying the Accuracy of Risk Management Models," *Journal of Derivatives* 3 (1995): 73–84.

should have a chi-square distribution with one degree of freedom. Values of the statistic are high for either very low or very high numbers of exceptions. There is a probability of 5% that the value of a chi-square variable with one degree of freedom will be greater than 3.84. It follows that we should reject the model whenever the expression in equation (11.11) is greater than 3.84.

Example 11.16

Suppose that, as in the previous two examples, we back-test a VaR model using 600 days of data when the VaR confidence level is 99%. The value of the statistic in equation (11.11) is greater than 3.84 when the number of exceptions, m , is one or less and when the number of exceptions is 12 or more. We therefore accept the VaR model when $2 \leq m \leq 11$ and reject it otherwise.

Generally speaking, the difficulty of back-testing a VaR model increases as the VaR confidence level increases. This is an argument in favor of using relatively low confidence levels for VaR for back-testing purposes and then using extreme value theory (see Chapter 12) to obtain the required confidence level.

11.10.1 Bunching

A separate issue from the number of exceptions is *bunching*. If daily portfolio changes are independent, exceptions should be spread evenly throughout the period used for back-testing. In practice, they are often bunched together, suggesting that losses on successive days are not independent. One approach to testing for bunching is to use the following statistic, suggested by Christoffersen (1998).⁶

$$-2 \ln[(1 - \pi)^{u_{00} + u_{10}} \pi^{u_{01} + u_{11}}] + 2 \ln[(1 - \pi_{01})^{u_{00}} \pi_{01}^{u_{01}} (1 - \pi_{11})^{u_{10}} \pi_{11}^{u_{11}}]$$

where u_{ij} is the number of observations in which we go from a day where we are in state i to a day where we are in state j . This statistic is chi-square with one degree of freedom if there is no bunching. State 0 is a day where there is no exception while state 1 is a day where there is an exception. Also,

$$\begin{aligned} \pi &= \frac{u_{01} + u_{11}}{u_{00} + u_{01} + u_{10} + u_{11}} \\ \pi_{01} &= \frac{u_{01}}{u_{00} + u_{01}} \\ \pi_{11} &= \frac{u_{11}}{u_{10} + u_{11}} \end{aligned}$$

⁶See P. F. Christoffersen, "Evaluating Interval Forecasts," *International Economic Review* 39 (1998): 841–862.

Summary

A value at risk (VaR) calculation is aimed at making a statement of the form: “We are X percent certain that we will not lose more than V dollars in time T .” The variable V is the VaR, X percent is the confidence level, and T is the time horizon. It has become a very popular risk measure. An alternative measure that provides better incentives for traders and has rather better theoretical properties is expected shortfall (ES). This is the expected loss conditional on the loss being greater than the VaR level. As Chapter 27 explains, regulators have switched from VaR to ES for some market risk capital calculations.

When changes in a portfolio value are normally distributed, it is easy to calculate VaR and ES from the mean and standard deviation of the change in the portfolio value during time T . If one-day changes in the value have zero-mean independent normal distributions, a T -day VaR (ES) equals the one-day VaR (ES) multiplied by $\sqrt{T}$. When the independence assumption is relaxed, other somewhat more complicated formulas can be used to go from the one-day VaR to the N -day VaR. In practice, losses often have heavier tails than the normal distribution. The power law is a way of modeling the tail of a distribution from empirical data. The theoretical basis for this approach is extreme value theory, which will be discussed in the next chapter.

Consider the situation where a portfolio has a number of subportfolios. The marginal value of a risk measure (VaR or ES) with respect to the i th subportfolio is the partial derivative of the risk measure with respect to the size of the subportfolio. The incremental VaR (ES) with respect to a particular subportfolio is the incremental effect of that subportfolio on VaR (ES). There is a formula that can be used for dividing VaR (ES) into components that correspond to the positions taken in the subportfolios. The component VaRs (ESs) sum to VaR (ES), and each component is, for a large portfolio of relatively small positions, approximately equal to the corresponding incremental VaR (ES).

Back-testing is an important activity. It examines how well a particular model for calculating a risk measure would have performed in the past. It is relatively easy to carry out for VaR. Back-testing may indicate weaknesses in a VaR model if the percentage of exceptions (that is, the percentage of times the actual loss exceeds VaR) is much greater or much less than that expected. There are statistical tests to determine whether a VaR model should be rejected because of the percentage of exceptions. As we will see in later chapters, regulators have rules for increasing market risk capital if they consider the results from back-testing over 250 days to be unsatisfactory.

Further Reading

- Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath. “Coherent Measures of Risk.” *Mathematical Finance* 9, (1999): 203–228.
- Basak, S., and A. Shapiro. “Value-at-Risk-Based Risk Management: Optimal Policies and Asset Prices.” *Review of Financial Studies* 14, no. 2 (2001): 371–405.
- Beder, T. “VaR: Seductive But Dangerous.” *Financial Analysts Journal* 51, no. 5 (1995): 12–24.

- Boudoukh, J., M. Richardson, and R. Whitelaw. "The Best of Both Worlds." *Risk* (May 1998): 64–67.
- Dowd, K. *Measuring Market Risk*. 2nd ed. Hoboken, NJ: John Wiley & Sons, 2005.
- Duffie, D., and J. Pan. "An Overview of Value at Risk." *Journal of Derivatives* 4, no. 3 (Spring 1997): 7–49.
- Hopper, G. "Value at Risk: A New Methodology for Measuring Portfolio Risk." *Business Review*, Federal Reserve Bank of Philadelphia (July–August 1996): 19–29.
- Hua, P., and P. Wilmot. "Crash Courses." *Risk* (June 1997): 64–67.
- Jackson, P., D. J. Maude, and W. Perraudin. "Bank Capital and Value at Risk." *Journal of Derivatives* 4, no. 3 (Spring 1997): 73–90.
- Jorion, P. *Value at Risk*. 3rd ed. New York: McGraw-Hill, 2006.
- Longin, F. M. "Beyond the VaR." *Journal of Derivatives* 8, no. 4 (Summer 2001): 36–48.
- Marshall, C., and M. Siegel. "Value at Risk: Implementing a Risk Measurement Standard." *Journal of Derivatives* 4, no. 3 (Spring 1997): 91–111.

Practice Questions and Problems (Answers at End of Book)

- 11.1 What is the difference between expected shortfall and VaR? What is the theoretical advantage of expected shortfall over VaR?
- 11.2 What conditions must be satisfied by the weights assigned to percentiles in a risk measure for the subadditivity condition in Section 11.5 to be satisfied?
- 11.3 A fund manager announces that the fund's one-month 95% VaR is 6% of the size of the portfolio being managed. You have an investment of \$100,000 in the fund. How do you interpret the portfolio manager's announcement?
- 11.4 A fund manager announces that the fund's one-month 95% expected shortfall is 6% of the size of the portfolio being managed. You have an investment of \$100,000 in the fund. How do you interpret the portfolio manager's announcement?
- 11.5 Suppose that each of two investments has a 0.9% chance of a loss of \$10 million and a 99.1% chance of a loss of \$1 million. The investments are independent of each other.
 - (a) What is the VaR for one of the investments when the confidence level is 99%?
 - (b) What is the expected shortfall for one of the investments when the confidence level is 99%?
 - (c) What is the VaR for a portfolio consisting of the two investments when the confidence level is 99%?
 - (d) What is the expected shortfall for a portfolio consisting of the two investments when the confidence level is 99%?
 - (e) Show that in this example VaR does not satisfy the subadditivity condition, whereas expected shortfall does.
- 11.6 Suppose that the change in the value of a portfolio over a one-day time period is normal with a mean of zero and a standard deviation of \$2 million; what is (a) the one-day 97.5% VaR, (b) the five-day 97.5% VaR, and (c) the five-day 99% VaR?

- 11.7 What difference does it make to your answer to Problem 11.6 if there is first-order daily autocorrelation with a correlation parameter equal to 0.16?
- 11.8 Explain carefully the differences between marginal VaR, incremental VaR, and component VaR for a portfolio consisting of a number of assets.
- 11.9 Suppose that we back-test a VaR model using 1,000 days of data. The VaR confidence level is 99% and we observe 17 exceptions. Should we reject the model at the 5% confidence level? Use a one-tailed test.
- 11.10 Explain what is meant by bunching.
- 11.11 Prove equation 11.5.
- 11.12 The change in the value of a portfolio in one month is normally distributed with a mean of zero and a standard deviation of \$2 million. Calculate the VaR and ES for a confidence level of 98% and a time horizon of three months.

Further Questions

- 11.13 Suppose that each of two investments has a 4% chance of a loss of \$10 million, a 2% chance of a loss of \$1 million, and a 94% chance of a profit of \$1 million. They are independent of each other.
- (a) What is the VaR for one of the investments when the confidence level is 95%?
 - (b) What is the expected shortfall when the confidence level is 95%?
 - (c) What is the VaR for a portfolio consisting of the two investments when the confidence level is 95%?
 - (d) What is the expected shortfall for a portfolio consisting of the two investments when the confidence level is 95%?
 - (e) Show that, in this example, VaR does not satisfy the subadditivity condition, whereas expected shortfall does.
- 11.14 Suppose the first-order autocorrelation for daily changes in the value of a portfolio is 0.12. The 10-day VaR, calculated by multiplying the one-day VaR by $\sqrt{10}$, is \$2 million. What is a better estimate of the VaR that takes account of autocorrelation?
- 11.15 Suppose that we back-test a VaR model using 1,000 days of data. The VaR confidence level is 99% and we observe 15 exceptions. Should we reject the model at the 5% confidence level? Use Kupiec's two-tailed test.
- 11.16 The change in the value of a portfolio in three months is normally distributed with a mean of \$500,000 and a standard deviation of \$3 million. Calculate the VaR and ES for a confidence level of 99.5% and a time horizon of three months.
- 11.17 The probability that the loss from a portfolio will be greater than \$10 million in one month is estimated to be 5%.
- (a) What is the one-month 99% VaR, assuming the change in value of the portfolio is normally distributed with zero mean?
 - (b) What is the one-month 99% VaR, assuming that the power law applies with $\alpha = 3$?

Chapter 12

Historical Simulation and Extreme Value Theory

In this chapter, we cover the most popular approach for calculating value at risk (VaR) and expected shortfall (ES) for market risk. It is known as historical simulation. It involves using the day-to-day changes in the values of market variables that have been observed in the past in a direct way to estimate the probability distribution of the change in the value of the current portfolio between today and tomorrow.

After describing the mechanics of the historical simulation approach, the chapter explains a number of extensions that can improve accuracy. It also covers stressed VaR and stressed ES, which are used by regulators to determine capital for market risk. Finally, it covers extreme value theory. This is a tool that can be used to improve VaR and ES estimates and to increase the confidence level used for them.

All the models covered in this chapter are illustrated with a portfolio consisting of an investment in four different stock indices. Historical data on the indices and VaR calculations can be found at www-2.rotman.utoronto.ca/~hull/RMFI/VaRExample.

12.1 The Methodology

Historical simulation involves using past data as a guide to what will happen in the future. Suppose that we want to calculate VaR for a portfolio using a one-day time horizon, a 99% confidence level, and 501 days of data. (We use 501 days of data because, as we will see, it leads to 500 scenarios being created.)

The first step is to identify the market variables affecting the portfolio. These market variables are sometimes referred to as *risk factors*. They typically include exchange rates, interest rates, stock indices, volatilities, and so on. Data are then collected on movements in these market variables over the most recent 501 days. This provides 500 alternative scenarios for what can happen between today and tomorrow. Denote the first day for which we have data as Day 0, the second day as Day 1, and so on. Scenario 1 is where the changes in the values of all variables are the same as they were between Day 0 and Day 1, Scenario 2 is where they are the same as between Day 1 and Day 2, and so on. For each scenario, the dollar change in the value of the portfolio between today and tomorrow is calculated. This defines a probability distribution for daily loss (with gains counted as negative losses) in the value of the portfolio. The 99 percentile of this distribution can be estimated as the fifth worst outcome.¹ The estimate of VaR is the loss when we are at this 99 percentile point. We are 99% certain that we will not take a loss greater than the VaR estimate if the changes in market variables in the past 500 days are representative of what will happen between today and tomorrow. ES is the average loss conditional that we are in the 1% tail of the loss distribution. VaR is estimated as the fifth worst loss. ES can be estimated by averaging the losses that are worse than VaR—that is, the four worst losses.²

To express the approach algebraically, define v_i as the value of a market variable on Day i and suppose that today is Day n . For most market variables it is appropriate to create scenarios by assuming that percentage changes will be repeated. The i th scenario in the historical simulation approach then assumes that the value of the market variable tomorrow will be

$$\text{Value under } i\text{th Scenario} = v_n \frac{v_i}{v_{i-1}} \quad (12.1)$$

For some variables such as interest rates, credit spreads, and volatilities, actual rather than percentage changes in market variables are considered. It is then the case that equation (12.1) becomes:

$$\text{Value under } i\text{th Scenario} = v_n + v_i - v_{i-1}$$

To simplify the discussion, in the rest of this chapter we will assume that the historical simulation is based on percentage changes in the underlying market variables.

12.1.1 Illustration

To illustrate the calculations underlying the approach, suppose that an investor in the United States owns, on July 8, 2020, a portfolio worth \$10 million consisting of

¹This is what is normally done, but there are alternatives. A case can be made for using the fifth worst loss, the sixth worst loss, or an average of the two. In Excel's PERCENTILE function, when there are n observations and k is an integer, the $k/(n-1)$ percentile is the observation ranked $k+1$. Other percentiles are calculated using linear interpolation.

²Footnote 1 pointed out that there are alternative VaR estimates that can be made from discrete data. The same is true for ES. We could, for example, average the five worst observations (i.e., include the VaR estimate). Alternatively, we could give the fifth worst loss half the weight of the others. What is suggested here seems to correspond to market practice.

investments in four stock indices: the S&P 500 in the United States, the FTSE 100 in the United Kingdom, the CAC 40 in France, and the Nikkei 225 in Japan. The value of the investment in each index on July 8, 2020, is shown in Table 12.1. An Excel spreadsheet containing 501 days of historical data on the closing prices of the four indices and a complete set of VaR calculations are on the author's website:³

www-2.rotman.utoronto.ca/~hull/RMFI/VaRExample

The calculations for this section are in worksheets 1 to 3. Note that the data we use are for *total return indices* where it is assumed that dividends are reinvested in the indices.

Because we are considering a U.S. investor, the values of the FTSE 100, CAC 40, and Nikkei 225 must be measured in U.S. dollars. For example, the FTSE 100 total return index stood at 6,515.12 on May 9, 2018 (the beginning of the 501-day period considered), when the exchange rate was 1.3553 USD per GBP. This means that, measured in U.S. dollars, it was at $6,515.12 \times 1.3553 = 8,830.23$. An extract from the data with all indices measured in U.S. dollars is shown in Table 12.2.

July 8, 2020 is an interesting date to choose in evaluating an equity investment. The COVID-19 pandemic had led to a big drop in equity markets in March 2020. By July

Table 12.1 Investment Portfolio Used for VaR Calculations July 8, 2020

Index	Portfolio Value (\$000s)
S&P 500	4,000
FTSE 100	3,000
CAC 40	1,000
Nikkei 225	2,000
Total	10,000

Table 12.2 U.S. Dollar Equivalent of Total Return Stock Indices for Historical Simulation Calculation

Day	Date	S&P 500	FTSE 100	CAC 40	Nikkei 225
0	May 9, 2018	5,292.90	8,830.23	16,910.33	322.40
1	May 10, 2018	5,343.70	8,926.56	16,915.41	321.24
2	May 11, 2018	5,354.69	8,982.76	17,065.64	326.20
3	May 14, 2018	5,359.66	8,999.31	17,121.67	328.03
...	...	...	...	...	...
499	July 7, 2020	6,445.59	7,269.36	15,784.97	345.40
500	July 8, 2020	6,496.14	7,255.04	15,540.44	342.01

³To keep the example as straightforward as possible, only days when all four indices traded were included in the compilation of the data. In practice, an attempt might be made to fill in data for days that were not U.S. holidays.

2020 the U.S market had largely recovered, but there was still a great deal of uncertainty about how long the pandemic would continue and when the economy would recover from it.

Table 12.3 shows the values of the indices (measured in U.S. dollars) on July 9, 2020, for the scenarios considered. Scenario 1 (the first row in Table 12.3) shows the values of indices on July 9, 2020, assuming that their percentage changes between July 8 and July 9, 2020, are the same as they were between May 9 and May 10, 2018; Scenario 2 (the second row in Table 12.3) shows the values of indices on July 9, 2020, assuming these percentage changes are the same as those between May 10 and May 11, 2018; and so on. In general, Scenario i assumes that the percentage changes in the indices between July 8 and July 9, 2020 are the same as they were between Day $i - 1$ and Day i for $1 \leq i \leq 500$. The 500 rows in Table 12.3 are the 500 scenarios considered.

The S&P 500 total return index was 6496.14 on July 8, 2020. On May 10, 2018, it was 5,343.70 up from 5,292.90 on May 9, 2018. The value of the S&P 500 under Scenario 1 is therefore

$$6,496.14 \times \frac{5,343.70}{5,292.90} = 6,558.49$$

Similarly, the value of the FTSE 100, the CAC 40, and the Nikkei 225 total return indices (measured in U.S. dollars) are 7,334.19, 15,545.12, and 340.79, respectively. The value of the portfolio under Scenario 1 is therefore (in \$000s):

$$\begin{aligned} &4,000 \times \frac{6,558.49}{6,496.14} + 3,000 \times \frac{7,334.19}{7,255.04} + 1,000 \times \frac{15,545.12}{15,540.44} + 2,000 \times \frac{340.79}{342.01} \\ &= 10,064.257 \end{aligned}$$

The portfolio therefore has a gain of \$64,257 under Scenario 1. A similar calculation is carried out for the other scenarios. A histogram for the losses (gains are negative losses) is shown in Figure 12.1. (The bars on the histogram represent losses, in \$000s, in the ranges 450 to 550, 350 to 450, 250 to 350 and so on.)

Table 12.3 Scenarios Generated for July 9, 2020, Using Data in Table 12.2 (negative losses are gains)

Scenario Number	S&P 500	FTSE 100	CAC 40	Nikkei 225	Portfolio Value (\$000s)	Loss (\$000s)
1	6,558.49	7,334.19	15,545.12	340.79	10,064.257	-64.257
2	6,509.50	7,300.72	15,678.46	347.28	10,066.822	-66.822
3	6,502.17	7,268.41	15,591.47	343.93	10,023.722	-23.722
...	...	...	...	...	...	...
...	...	...	...	...	...	...
499	6,425.90	7,293.40	15,543.89	341.21	9,968.126	31.874
500	6,547.09	7,240.75	15,299.71	338.66	9,990.361	9.639

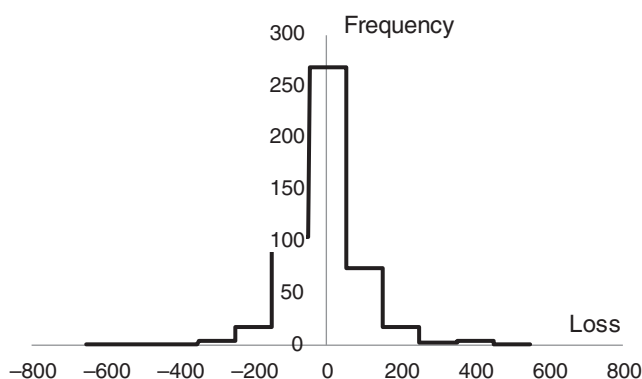


Figure 12.1 Histogram of Losses between July 8, 2020, and July 9, 2020 for the 500 Scenarios Considered (Gains are negative losses.)

Table 12.4 Losses Ranked from Highest to Lowest for 500 Scenarios

Scenario Number	Loss (\$000s)
427	922.484
429	858.423
424	653.541
415	490.215
482	422.291
440	362.733
426	360.532
431	353.788
417	323.505
433	305.216
452	245.151
418	241.561
140	231.269
289	230.626
152	229.683
...	...
...	...
...	...

The losses for the 500 different scenarios are then ranked. An extract from the results of doing this is shown in Table 12.4. The worst scenario is number 427. The one-day 99% value at risk can be estimated as the fifth worst loss. This is \$422,291.

Regulators have in the past used the 10-day VaR for determining market risk capital with the understanding that this can be calculated as $\sqrt{10}$ times the one-day 99% VaR. (See Section 11.6.1.) In our example this would give the 10-day VaR as

$$\sqrt{10} \times \$422,291 = \$1,335,401$$

As we will see in Chapter 27, regulators now calculate capital (when using the internal models approach) from a series of ten-day expected shortfall calculations where overlapping 10-day periods are used.

Each day, the VaR estimate in our example would be updated using the most recent 501 days of data. Consider, for example, what happens on July 9, 2020 (Day 501). We find out new values for all the market variables and are able to calculate a new value for our portfolio. We then go through the procedure we have outlined to calculate a new VaR. Data on the market variables from May 10, 2018, to July 9, 2020 (Day 1 to Day 501), are used in the calculation. (This gives us the required 500 observations on the percentage changes in market variables; the May 9, 2018, Day 0, values of the market variables are no longer used.) Similarly, on the next trading day, July 10, 2020 (Day 502), data from May 11, 2018, to July 10, 2020 (Day 2 to Day 502), are used to determine VaR; and so on.

In practice, a financial institution's portfolio is, of course, considerably more complicated than the one we have considered here. It is likely to consist of thousands or tens of thousands of positions. Often some of the positions are in forward contracts, options, and other derivatives, which have to be valued for each of the scenarios considered. The portfolio itself is likely to change from day to day. If the financial institution's trading leads to a riskier portfolio, the 10-day 99% VaR typically increases; if it leads to a less risky portfolio, the VaR typically decreases. The VaR on any given day is calculated on the assumption that the portfolio will remain unchanged over the next business day.

The market variables (or risk factors) that have to be considered in a VaR calculation include exchange rates, commodity prices, and interest rates. In the case of interest rates, a financial institution typically needs several term structures describing different zero-coupon interest rates in each of a number of different currencies in order to value its portfolio. The market variables that are considered are the ones from which these term structures are calculated (see Appendix B for the calculations to obtain the zero-coupon term structure of interest rates). There might be as many as 10 market variables for each zero curve to which the financial institution is exposed.

12.1.2 Expected Shortfall

To calculate expected shortfall using historical simulation, we average the losses that are worse than VaR. In the case of our example, the four worst losses (\$000s) are from scenarios 427, 429, 424, and 415 (see Table 12.4). The average of the losses for these scenarios is \$731,166. This is the expected shortfall estimate. (It is debatable whether the average should be taken over the five worst scenarios rather than the four worst ones. Regulators when they use expected shortfall typically only consider scenarios that give rise to losses that are greater than the VaR level.)

12.1.3 Stressed VaR and Stressed ES

The calculations given so far assume that the most recent data are used for the historical simulation on any given day. For example, when calculating VaR and ES for the four-index example we used data from the immediately preceding 501 days. We will refer to VaR and ES calculated in this way as *current VaR* and *current ES*. However, historical simulations can be based on data from any period in the past. Periods of high volatility will tend to give high values for VaR and ES, whereas periods of low volatility will tend to give low values.

Regulators have introduced measures known as *stressed VaR* and *stressed ES*. To calculate the measures, a financial institution must search for the 251-day period that would be particularly stressful for its current portfolio. The data for that 251-day period then play the same role as the 501-day period in our example. The changes in market variables between Day 0 and Day 1 of the 251-day period are used to create the first scenario; the changes in market variables between Day 1 and Day 2 of the 251-day period are used to create the second scenario; and so on. In total, 250 scenarios are created. The one-day 99% stressed VaR can be calculated as the loss that is midway between the loss for the second worst scenario and the loss for the third worst scenario. The one-day 99% ES can be calculated as the average of the worst two losses.⁴

12.2 Accuracy of VaR

The historical simulation approach estimates the distribution of portfolio changes from a finite number of observations. As a result, the estimates of percentiles of the distribution are subject to error.

Kendall and Stuart (1972) describe how to calculate a confidence interval for the percentile of a probability distribution when it is estimated from sample data.⁵ Suppose that the q -percentile of the distribution is estimated as x . The standard error of the estimate is

$$\frac{1}{f(x)} \sqrt{\frac{(1-q)q}{n}}$$

where n is the number of observations and $f(x)$ is an estimate of the probability density function of the loss evaluated at x . The probability density, $f(x)$, can be estimated approximately by fitting the empirical data to an appropriate distribution whose properties are known.

⁴Again we should mention that there are alternative approaches. VaR could be calculated as the third worst loss or the second worst loss. (Some regulators such as the Federal Reserve in the United States prefer the latter.) A case can be made for calculating ES as $0.4c_1 + 0.4c_2 + 0.2c_3$ where c_1 , c_2 , and c_3 are the three worst losses with $c_1 > c_2 > c_3$ although, as indicated earlier, regulators prefer the more conservative approach of averaging c_1 and c_2 .

⁵See M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, vol. 1, *Distribution Theory*, 4th ed. (London: Charles Griffin, 1972).

Example 12.1

Suppose we are interested in estimating the 99th percentile of a loss distribution from 500 observations so that $n = 500$ and $q = 0.99$. We can estimate $f(x)$ by approximating the actual empirical distribution with a standard distribution whose properties are known. Suppose that a normal distribution is chosen as the standard distribution and the best-fit mean and standard deviation are zero and \$10 million, respectively. Using Excel, the 99th percentile is `NORMINV(0.99,0,10)` or 23.26. The value of $f(x)$ is `NORMDIST(23.26,0,10,FALSE)` or 0.0027. The standard error of the estimate that is made is

$$\frac{1}{0.0027} \times \sqrt{\frac{0.01 \times 0.99}{500}} = 1.67$$

If the estimate of the 99th percentile using historical simulation is \$25 million, a 95% confidence interval is from $25 - 1.96 \times 1.67$ to $25 + 1.96 \times 1.67$, that is, from \$21.7 million to \$28.3 million.

As Example 12.1 illustrates, the standard error of a VaR estimated using historical simulation tends to be quite high. It decreases as the VaR confidence level is decreased. For example, if in Example 12.1 the VaR confidence level had been 95% instead of 99%, the standard error would be \$0.95 million instead of \$1.67 million. The standard error declines as the sample size is increased—but only as the square root of the sample size. If we quadrupled the sample size in Example 12.1 from 500 to 2,000 observations, the standard error halves from \$1.67 million to about \$0.83 million.

Additionally, we should bear in mind that historical simulation assumes that the joint distribution of daily changes in market variables is stationary through time. This is unlikely to be exactly true and creates additional uncertainty about VaR.

In the case of the data considered in Tables 12.1 to 12.4, when the loss is measured in \$000s, the mean is -1.660 and the standard deviation is 120.146. If a normal distribution is assumed, a similar calculation to that in Example 12.1 gives $f(x)$ as 0.000222 and the standard error of the estimate (in \$000s) is

$$\frac{1}{0.000222} \times \sqrt{\frac{0.01 \times 0.99}{500}} = 20.059$$

The estimate of VaR is \$422,291. This shows that a 95% confidence interval for the VaR is about \$380,000 to \$460,000.

The normal distribution is not a particularly good assumption for the loss distribution, because losses have heavier tails than the normal distribution. Better standard error estimates can be obtained by assuming a Pareto distribution for $f(x)$, as discussed in Section 12.5.

12.3 Extensions

When current VaR and current ES are calculated (using data from an immediately preceding period), the assumption is that recent history is in some sense a good guide to the future. More precisely, it is that the empirical probability distribution estimated for market variables over the immediately preceding period is a good guide to the behavior of the market variables over the next day. Unfortunately, the behavior of market variables is nonstationary. Sometimes the volatility of a market variable is high; sometimes it is low. In this section, we cover extensions of the basic historical simulation approach in Section 12.1 that are designed to adjust for nonstationarity. We also show how an approach known as the bootstrap method can be used to determine standard errors.

12.3.1 *Weighting of Observations*

The basic historical simulation approach assumes that each day in the past is given equal weight. More formally, if we have observations for n day-to-day changes, each of them is given a weighting of $1/n$. Boudoukh, Richardson, and Whitelaw (1998) suggest that more recent observations should be given more weight because they are more reflective of current volatilities and current macroeconomic conditions.⁶ The natural weighting scheme to use is one where weights decline exponentially. (We used this when developing the exponentially weighted moving average model for monitoring volatility in Chapter 8.) The weight assigned to Scenario 1 (which is the one calculated from the most distant data) is λ times that assigned to Scenario 2. This in turn is λ times that given to Scenario 3, and so on. So that the weights add up to 1, the weight given to Scenario i is

$$\frac{\lambda^{n-i}(1 - \lambda)}{1 - \lambda^n}$$

where n is the number of scenarios. As λ approaches 1, this approaches the basic historical simulation approach where all observations are given a weight of $1/n$. (See Problem 12.2.)

VaR is calculated by ranking the observations from the worst outcome to the best. Starting at the worst outcome, weights are summed until the required percentile of the distribution is reached. For example, if we are calculating VaR with a 99% confidence level, we continue summing weights until the sum just exceeds 0.01. We have then reached the 99% VaR level. The parameter λ can be chosen by trying different values and seeing which one back-tests best. One disadvantage of the exponential weighting approach relative to the basic historical simulation approach is that the effective sample size is reduced. However, we can compensate for this by using a larger value of n . Indeed, it is not really necessary to discard old days as we move forward in time, because they are given relatively little weight.

⁶See J. Boudoukh, M. Richardson, and R. Whitelaw, "The Best of Both Worlds: A Hybrid Approach to Calculating Value at Risk," *Risk* 11 (May 1998): 64–67.

Table 12.5 Losses Ranked from Highest to Lowest for 500 Scenarios, with Weights

Scenario Number	Loss (\$000s)	Weight	Cumulative Weight
427	922.484	0.00378	0.00378
429	858.423	0.00381	0.00759
424	653.541	0.00372	0.01131
415	490.215	0.00356	0.01486
482	422.292	0.00497	0.01984
440	362.733	0.00403	0.02387
426	360.532	0.00376	0.02763
431	353.788	0.00385	0.03148
417	323.505	0.00359	0.03507
...	...	...	...
...	...	...	...
...	...	...	...

Table 12.5 shows the results of using this procedure for the portfolio considered in Section 12.1 with $\lambda = 0.995$. (See worksheets 4 and 5 on the author's website.) The value of VaR when the confidence level is 99% is now the third worst loss, \$653,541 (not the fifth worst loss of \$422,291). The reason for this result is that recent observations are given more weight and the largest losses have occurred relatively recently. The standard calculation in Section 12.1 gives all observations a weighting of $1/500 = 0.002$. The highest loss occurred on Scenario 427, and this scenario has a weight of

$$\frac{(0.995^{73}) \times 0.005}{1 - 0.995^{500}} = 0.00378$$

The 0.01 tail of the loss distribution consists of a probability 0.00378 of a loss of \$922,484, a 0.00381 probability of a loss of \$858,423, and a $0.01 - 0.00378 - 0.00381 = 0.00241$ probability of a loss of \$653,541. The expected shortfall can therefore be calculated as⁷

$$\frac{0.00378 \times \$922,484 + 0.00381 \times \$858,423 + 0.00241 \times \$653,541}{0.01} = \$833,228$$

12.3.2 Volatility Scaling for Market Variables

Hull and White (1998) suggest a way of incorporating estimates of volatility into the historical simulation approach.⁸ Define the daily volatility for a particular market variable estimated at the end of day $i - 1$ as σ_i . (Estimation procedures have been discussed in Chapter 8.) This can be considered to be an estimate of the daily volatility between the end of day $i - 1$ and the end of day i . Suppose that it is now day n so that the current

⁷This calculation seems to fit in with the nature of weighted observations approach best. An alternative would be to calculate a weighted average of the two worst losses.

⁸See J. Hull and A. White, "Incorporating Volatility Updating into the Historical Simulation Method for Value at Risk," *Journal of Risk* (Fall 1998): 5–19.

estimate of the volatility of the market variable (i.e., the volatility between today and tomorrow) is σ_{n+1} .

Suppose that σ_{n+1} is twice σ_i for a particular market variable. This means that we estimate the daily volatility of the market variable to be twice as great today as on day $i - 1$. The changes we expect to see between today and tomorrow are twice as big as changes between day $i - 1$ and day i . When carrying out the historical simulation and creating a sample of what could happen between today and tomorrow based on what happened between day $i - 1$ and day i , it therefore makes sense to multiply the latter by 2. In general, when this approach is used, the expression in equation (12.1) for the value of a market variable under the i th scenario becomes

$$\text{Value under } i\text{th Scenario} = v_n \frac{v_{i-1} + (v_i - v_{i-1})\sigma_{n+1}/\sigma_i}{v_{i-1}} \quad (12.2)$$

Each market variable can be handled in the same way.

This approach takes account of volatility changes in a natural and intuitive way and produces VaR estimates that incorporate more current information. The VaR estimates can be greater than any of the historical losses that would have occurred for the current portfolio during the historical period considered. Hull and White produce evidence using exchange rates and stock indices to show that this approach is superior to traditional historical simulation and to the exponential weighting scheme described earlier.

For the data in Table 12.2, the daily volatility estimates, calculated using the exponentially weighted moving average (EWMA) method with the λ parameter equal to 0.94, are shown in Table 12.6.⁹ (See worksheets 6 to 8 on the author's website.) The ratios of the volatility estimated for July 9, 2020 (last row of table), to the volatility estimated for May 10, 2018 (first row of table), are 1.06, 1.21, 1.23, and 1.25 for the S&P 500, FTSE 100, CAC 40, and Nikkei 225, respectively. These are used as multipliers for the actual changes in the indices between May 9 and May 10, as indicated in equation (12.2). Similarly, the ratios of the volatility estimated for July 9, 2020 (last row of Table 12.6),

Table 12.6 Volatilities (% per Day) Estimated for the Following Day Using EWMA

Day	Date	S&P 500	FTSE 100	CAC 40	Nikkei 225
0	May 9, 2018	1.66	1.37	1.54	1.32
1	May 10, 2018	1.63	1.35	1.49	1.28
2	May 11, 2018	1.58	1.32	1.46	1.30
3	May 14, 2018	1.53	1.28	1.42	1.26
...	...	...	...	...	...
...	...	...	...	...	...
499	July 7, 2020	1.80	1.70	1.91	1.68
500	July 8, 2020	1.76	1.65	1.89	1.65

⁹A decision must be made on how to start the variance time series. The initial variance in the calculations reported here was the sample variance calculated over the whole time period.

to the volatility estimated for May 11, 2018 (second row of Table 12.6), are 1.08, 1.22, 1.27, and 1.29 for the S&P 500, FTSE 100, CAC 40, and Nikkei 225, respectively. These are used as multipliers for the actual changes in the indices between May 10 and May 11, 2018. Multipliers for the other 498 daily changes are calculated in the same way.

Table 12.7 shows an extract from a table that ranks losses from the highest to the lowest. Comparing this with Table 12.4, we see that the scenarios generating the biggest losses are not the same. This is because a moderate loss when the estimated volatility is low can lead to a more extreme scenario than a bigger loss when the estimated volatility is high. The one-day 99% VaR is \$566,492. The one-day ES (calculated from the 4 worst scenarios) is \$705,855. These are more than twice as high as the estimates given by standard calculations.

12.3.3 Volatility Scaling for the Portfolio

A variation on the approach we just have described is to use EWMA to monitor the standard deviation of the simulated losses given by successive scenarios in the standard approach in Section 12.1. The losses are those given in the final column of Table 12.3. (See worksheets 9 and 10 on the author’s website for the calculations.) An adjusted loss for the i th scenario is then calculated by multiplying the loss given by the standard approach by the ratio of the estimated standard deviation for the last (500th) scenario to the estimated standard deviation for the i th scenario. This procedure is much simpler than incorporating volatility on a variable-by-variable basis and has the advantage that changing correlations as well as changing volatilities are implicitly considered. Table 12.8 shows the calculation of the portfolio loss standard deviation and the adjusted losses for our example. (Similarly to before, the Day 1 loss variance is the sample variance calculated over the whole 500-day period, and the λ parameter used in EWMA calculations is 0.94.)

Table 12.7 Losses Ranked from Highest to Lowest for 500 Scenarios When Market Variable Movements Are Adjusted for Volatility

Scenario Number	Loss (\$000s)
415	1,064.165
427	607.359
102	583.832
424	568.064
289	566.492
397	502.621
103	437.660
140	430.578
326	426.327
...	...
...	...
...	...

Table 12.8 Results When Volatility of Simulated Losses Is Monitored

Scenario Number	Loss Given by Standard Approach	Loss SD (\$000s)	SD Ratio	Adjusted Loss
1	−64.257	120.146	1.116	−71.708
2	−66.822	117.544	1.141	−76.220
3	−23.762	115.133	1.165	−27.672
...	...	...	...	...
...	...	...	...	...
499	31.874	138.056	0.971	30.955
500	9.639	134.078	1.000	9.639

Table 12.9 Adjusted Losses Ranked from Highest to Lowest for 500 Scenarios When Simulated Losses Are Adjusted for Their Volatility

Scenario Number	Adjusted Loss (\$000s)
405	1005.299
397	545.326
289	532.875
424	517.985
427	517.195
326	461.734
103	452.468
288	400.709
102	380.790
...	...
...	...
...	...

Table 12.9 shows the ranked adjusted losses. The one-day 99% VaR is \$517,195, and the one-day expected shortfall is \$650,354.

12.3.4 Bootstrap Method

The bootstrap method is a variation on the basic historical simulation approach, aimed at calculating a confidence interval for VaR (current or stressed).¹⁰ It involves creating a set of changes in the portfolio value based on historical movements in market variables in the usual way. We then sample with replacement from these changes to create many new similar data sets. We calculate the VaR for each of the new data sets. Our 95% confidence interval for VaR is the range between the 2.5 percentile point and the 97.5 percentile point of the distribution of the VaRs calculated from the data sets.

¹⁰See P. Christoffersen and S. Gonçalves, “Estimation Risk in Financial Risk Management,” *Journal of Risk* 7, no. 3 (2007): 1–28.

Suppose, for example, that we have 500 days of data. We could sample with replacement 500,000 times from the data to obtain 1,000 different sets of 500 days of data. We calculate the VaR for each set. We then rank the VaRs. Suppose that the 25th largest VaR is \$5.3 million and the 975th largest VaR is \$8.9 million. The 95% confidence interval for VaR is \$5.3 million to \$8.9 million. Usually, the width of the confidence interval calculated for VaR using the bootstrap method is less than that calculated using the procedure in Section 12.2.

12.4 Computational Issues

Historical simulation involves valuing the whole portfolio of a financial institution many times (500 times in our example). This can be computationally very time consuming. This is particularly true when some of the instruments in the portfolio are valued with Monte Carlo simulation, because there is then a simulation within a simulation problem because each trial of the historical simulation involves a Monte Carlo simulation.

To reduce computation time, financial institutions sometimes use an approximation. Consider an instrument whose price, P , is dependent on a single market variable, S . An approximate estimate of the change, ΔP , in P resulting from a change, ΔS , in S is given by a Taylor series expansion:

$$\Delta P = \delta \Delta S + \frac{1}{2} \gamma (\Delta S)^2 \quad (12.3)$$

where δ and γ are the first and second partial derivatives of P with respect to S . The values of δ and γ are usually known because they are calculated for hedging purposes. (See Chapter 15 for a discussion of this.) This equation can therefore be used as a fast approximate way of calculating the changes in the value of the transaction for the changes in the value of S that are considered by the historical simulation.

When an instrument depends on several market variables, $S_i (1 \leq i \leq n)$, equation (12.3) becomes

$$\Delta P = \sum_{i=1}^n \delta_i \Delta S_i + \sum_{i=1}^n \sum_{j=1}^n \frac{1}{2} \gamma_{ij} \Delta S_i \Delta S_j \quad (12.4)$$

where δ_i and γ_{ij} are defined as

$$\delta_i = \frac{\partial P}{\partial S_i} \quad \gamma_{ij} = \frac{\partial^2 P}{\partial S_i \partial S_j}$$

12.5 Extreme Value Theory

Section 8.4 introduced the power law and explained that it can be used to estimate the tails of a wide range of distributions. We now provide the theoretical underpinnings for the power law and present estimation procedures that are more sophisticated than

those used in Section 8.4. *Extreme value theory* (EVT) is the term used to describe the science of estimating the tails of a distribution. EVT can be used to improve VaR or ES estimates and to help in situations where analysts want to estimate VaR with a very high confidence level. It is a way of smoothing and extrapolating the tails of an empirical distribution.

12.5.1 The Key Result

The key result in EVT was proved by Gnedenko (1943).¹¹ It shows that the tails of a wide range of different probability distributions share common properties.

Suppose that $F(v)$ is the cumulative distribution function for a variable v (such as the loss on a portfolio over a certain period of time) and that u is a value of v in the right-hand tail of the distribution. The probability that v lies between u and $u + \gamma$ ($\gamma > 0$) is $F(u + \gamma) - F(u)$. The probability that v is greater than u is $1 - F(u)$. Define $F_u(\gamma)$ as the probability that v lies between u and $u + \gamma$ conditional on $v > u$. This is

$$F_u(\gamma) = \frac{F(u + \gamma) - F(u)}{1 - F(u)}$$

The variable $F_u(\gamma)$ defines the right tail of the probability distribution. It is the cumulative probability distribution for the amount by which v exceeds u given that it does exceed u .

Gnedenko's result states that, for a wide class of distributions $F(v)$, the distribution of $F_u(\gamma)$ converges to a generalized Pareto distribution as the threshold u is increased. The generalized Pareto (cumulative) distribution is

$$G_{\xi, \beta}(\gamma) = 1 - \left[1 + \xi \frac{\gamma}{\beta} \right]^{-1/\xi} \quad (12.5)$$

The distribution has two parameters that have to be estimated from the data. These are ξ and β . The parameter ξ is the shape parameter and determines the heaviness of the tail of the distribution. The parameter β is a scale parameter.

When the underlying variable v has a normal distribution, $\xi = 0$.¹² As the tails of the distribution become heavier, the value of ξ increases. For most financial data, ξ is positive and in the range 0.1 to 0.4.¹³

¹¹See D. V. Gnedenko, "Sur la distribution limitée du terme d'une série aléatoire," *Annals of Mathematics* 44 (1943): 423–453.

¹²When $\xi = 0$, the generalized Pareto distribution becomes

$$G_{\xi, \beta}(\gamma) = 1 - \exp\left(-\frac{\gamma}{\beta}\right)$$

¹³One of the properties of the distribution in equation (12.5) is that the k th moment of v , $E(v^k)$, is infinite for $k \geq 1/\xi$. For a normal distribution, all moments are finite. When $\xi = 0.25$, only the first three moments are finite; when $\xi = 0.5$, only the first moment is finite; and so on.

12.5.2 Estimating ξ and β

The parameters ξ and β can be estimated using maximum likelihood methods (see Section 8.9 for a discussion of maximum likelihood methods). The probability density function, $g_{\xi,\beta}(\gamma)$, of the cumulative distribution in equation (12.5) is calculated by differentiating $G_{\xi,\beta}(\gamma)$ with respect to γ . It is

$$g_{\xi,\beta}(\gamma) = \frac{1}{\beta} \left(1 + \frac{\xi\gamma}{\beta} \right)^{-1/\xi-1} \quad (12.6)$$

We first choose a value for u . (A value close to the 95th percentile point of the empirical distribution usually works well.) We then rank the observations on ν from the highest to the lowest and focus our attention on those observations for which $\nu > u$. Suppose there are n_u such observations and they are ν_i ($1 \leq i \leq n_u$). From equation (12.6), the likelihood function (assuming that $\xi \neq 0$) is

$$\prod_{i=1}^{n_u} \frac{1}{\beta} \left(1 + \frac{\xi(\nu_i - u)}{\beta} \right)^{-1/\xi-1}$$

Maximizing this function is the same as maximizing its logarithm:

$$\sum_{i=1}^{n_u} \ln \left[\frac{1}{\beta} \left(1 + \frac{\xi(\nu_i - u)}{\beta} \right)^{-1/\xi-1} \right] \quad (12.7)$$

Standard numerical procedures can be used to find the values of ξ and β that maximize this expression. Excel's Solver produces good results.

12.5.3 Estimating the Tail of the Distribution

The probability that $\nu > u + \gamma$ conditional that $\nu > u$ is $1 - G_{\xi,\beta}(\gamma)$. The probability that $\nu > u$ is $1 - F(u)$. The unconditional probability that $\nu > x$ (when $x > u$) is therefore

$$[1 - F(u)][1 - G_{\xi,\beta}(x - u)]$$

If n is the total number of observations, an estimate of $1 - F(u)$, calculated from the empirical data, is n_u/n . The unconditional probability that $\nu > x$ is therefore

$$\text{Prob}(\nu > x) = \frac{n_u}{n} [1 - G_{\xi,\beta}(x - u)] = \frac{n_u}{n} \left[1 + \xi \frac{x - u}{\beta} \right]^{-1/\xi} \quad (12.8)$$

12.5.4 Equivalence to the Power Law

If we set $u = \beta/\xi$, equation (12.8) reduces to

$$\text{Prob}(\nu > x) = \frac{n_u}{n} \left[\frac{\xi x}{\beta} \right]^{-1/\xi}$$

This is

$$Kx^{-\alpha}$$

where

$$K = \frac{n_u}{n} \left[\frac{\xi}{\beta} \right]^{-1/\xi}$$

and $\alpha = 1/\xi$. This shows that equation (12.8) is consistent with the power law introduced in Section 8.4.

12.5.5 The Left Tail

The analysis so far has assumed that we are interested in the right tail of the probability distribution of a variable ν . If we are interested in the left tail of the probability distribution, we can work with $-\nu$ instead of ν . Suppose, for example, that an oil company has collected data on daily percentage increases in the price of oil and wants to estimate a VaR that is the one-day percentage decline in the price of oil that has a 99.9% probability of not being exceeded. This is a statistic calculated from the left tail of the probability distribution of oil price increases. The oil company would change the sign of each data item (so that the data were measuring oil price decreases rather than increases) and then use the methodology that has been presented.

12.5.6 Calculation of VaR and ES

To calculate VaR with a confidence level of q , it is necessary to solve the equation

$$F(\text{VaR}) = q$$

Because $F(x) = 1 - \text{Prob}(\nu > x)$, equation (12.8) gives

$$q = 1 - \frac{n_u}{n} \left[1 + \xi \frac{\text{VaR} - u}{\beta} \right]^{-1/\xi}$$

so that

$$\text{VaR} = u + \frac{\beta}{\xi} \left\{ \left[\frac{n}{n_u} (1 - q) \right]^{-\xi} - 1 \right\} \quad (12.9)$$

The expected shortfall is given by

$$\text{ES} = \frac{\text{VaR} + \beta - \xi u}{1 - \xi} \quad (12.10)$$

12.6 Applications of EVT

Consider again the data in Tables 12.1 to 12.4. When $u = 160$, $n_u = 25$ (that is, there are 25 scenarios where the loss in \$000s is greater than 160). Table 12.10 shows calculations

Table 12.10 Extreme Value Theory Calculations for Table 12.4 (the parameter u is 160 and trial values for β and ξ are 40 and 0.3, respectively)

Scenario Number	Loss (\$000s)	Rank	$\ln \left[\frac{1}{\beta} \left(1 + \frac{\xi(v_i - u)}{\beta} \right)^{-1/\xi - 1} \right]$
427	922.484	1	-11.94
429	858.423	2	-11.62
424	653.541	3	-10.40
415	490.215	4	-9.09
482	422.291	5	-8.40
440	362.733	6	-7.69
426	360.532	7	-7.67
431	353.788	8	-7.58
...	...	...	...
...	...	...	...
...	...	...	...
150	163.620	25	-3.80
			-160.99
Trial Estimates of EVT Parameters			
ξ	β		
0.3	40		

for the trial values $\beta = 40$ and $\xi = 0.3$. The value of the log-likelihood function in equation (12.7) is -160.99 .

When Excel's Solver is used to search for the values of β and ξ that maximize the log-likelihood function (see worksheet 11 on the author's website), it gives

$$\begin{aligned}\beta &= 110.46 \\ \xi &= 0.354\end{aligned}$$

and the maximum value of the log-likelihood function is -151.47 .

Suppose that we wish to estimate the probability that the portfolio loss between July 8 and July 9, 2020, will be more than \$300,000 (or 3% of its value). From equation (12.8) this is

$$\frac{25}{500} \left[1 + 0.354 \frac{300 - 160}{110.46} \right]^{-1/0.3545} = 0.0176$$

This is more accurate than counting observations. The probability that the portfolio loss will be more than \$500,000 (or 5% of its value) is similarly 0.0062.

From equation (12.9), the value of VaR with a 99% confidence limit is

$$160 + \frac{110.46}{0.354} \left\{ \left[\frac{500}{25} (1 - 0.99) \right]^{-0.354} - 1 \right\} = 399.6$$

or \$399,600. (In this instance, the VaR estimate is about \$23,000 less than the fifth worst loss.) When the confidence level is increased to 99.9%, VaR becomes

$$160 + \frac{110.46}{0.354} \left\{ \left[\frac{500}{25} (1 - 0.999) \right]^{-0.354} - 1 \right\} = 1094.6$$

or \$1,094,600. When it is increased further to 99.97%, VaR becomes

$$160 + \frac{110.46}{0.354} \left\{ \left[\frac{500}{25} (1 - 0.9997) \right]^{-0.354} - 1 \right\} = 1757.4$$

or \$1,757,400.

The formula in equation (12.10) can improve ES estimates and allow the confidence level used for ES estimates to be increased. In our example, when the confidence level is 99%, the estimated ES is

$$\frac{399.6 + 110.46 - 0.354 \times 160}{1 - 0.354} = 702.0$$

or \$702,000. When the confidence level is 99.9%, the estimated ES is

$$\frac{1094.6 + 110.46 - 0.354 \times 160}{1 - 0.354} = 1778.1$$

or \$1,778,100.

EVT can be used for current or stressed measures. It can also be used in a straightforward way in conjunction with the volatility-scaling procedures in Section 12.3 (see Problem 12.11). It can also be used in conjunction with the weighting-of-observations procedure in Section 12.3. In this case, the terms being summed in equation (12.7) must be multiplied by the weights applicable to the underlying observations.

A final calculation can be used to refine the confidence interval for the 99% VaR estimate in Section 12.2. The probability density function evaluated at the VaR level for the probability distribution of the loss, conditional on it being greater than 160, is given by the $g_{\xi, \beta}$ function in equation (12.6). It is

$$\frac{1}{110.5} \left(1 + \frac{0.354 \times (399.6 - 160)}{110.5} \right)^{-1/0.354-1} = 0.0001024$$

The unconditional probability density function evaluated at the VaR level is $n_u/n = 25/500$ times this or 0.00005. Not surprisingly, this is lower than the 0.000222 estimated in Section 12.2 and leads to a wider confidence interval for VaR.

12.6.1 Choice of u

A natural question is: “How do the results depend on the choice of u ?” It is often found that values of ξ and β do depend on u , but the estimates of $F(x)$ remain roughly the same. (Problem 12.10 considers what happens when u is changed from 160 to 150 in the example we have been considering.) We want u to be sufficiently high that we are

truly investigating the shape of the tail of the distribution, but sufficiently low that the number of data items included in the maximum likelihood calculation is not too low. More data lead to more accuracy in the assessment of the shape of the tail. We have applied the procedure with 500 data items. Ideally, more data would be used.

A rule of thumb is that u should be approximately equal to the 95th percentile of the empirical distribution. In the search for the optimal values of ξ and β , both variables should be constrained to be positive. If the optimizer tries to set ξ negative, it is likely to be a sign that either (a) the tail of the distribution is not heavier than the normal distribution or (b) an inappropriate value of u has been chosen.

Summary

Historical simulation is a very popular approach for estimating VaR or ES. It involves creating a database consisting of the daily movements in all market variables over a period of time. The first simulation trial assumes that the percentage change in each market variable is the same as that on the first day covered by the database, the second simulation trial assumes that the percentage changes are the same as those on the second day, and so on. The change in the portfolio value is calculated for each simulation trial, and VaR is calculated as the appropriate percentile of the probability distribution of this change, while ES is calculated as the average change in value in the tail of this distribution.

The standard error for a VaR that is estimated using historical simulation tends to be quite high. The higher the VaR confidence level required, the higher the standard error. In one extension of the basic historical simulation approach, the weights given to observations decrease exponentially as the observations become older; in another, adjustments are made to historical data to reflect changes in volatility.

Extreme value theory is a way of smoothing the tails of the probability distribution of portfolio daily changes calculated using historical simulation. It avoids the calculation uncertainties mentioned in footnotes 1, 2, 4, and 7. It leads to estimates of VaR and ES that reflect the whole shape of the tail of the distribution, not just the positions of a few losses in the tails. Extreme value theory can also be used to estimate VaR and ES when the confidence level is very high. For example, even if we have only 500 days of data, it could be used to come up with an estimate of VaR or ES for a confidence level of 99.9%.

Further Reading

- Boudoukh, J., M. Richardson, and R. Whitelaw. "The Best of Both Worlds." *Risk* (May 1998): 64–67.
- Embrechts, P., C. Kluppelberg, and T. Mikosch. *Modeling Extremal Events for Insurance and Finance*. New York: Springer, 1997.
- Hendricks, D. "Evaluation of Value-at-Risk Models Using Historical Data," *Economic Policy Review*, Federal Reserve Bank of New York, vol. 2 (April 1996): 39–69.

- Hull, J. C., and A. White. "Incorporating Volatility Updating into the Historical Simulation Method for Value at Risk." *Journal of Risk* 1, no. 1 (1998): 5–19.
- McNeil, A. J. "Extreme Value Theory for Risk Managers." In *Internal Modeling and CAD II* (London: Risk Books, 1999). See also www.macs.hw.ac.uk/~mcneil/ftp/cad.pdf.
- Neftci, S. N. "Value at Risk Calculations, Extreme Events and Tail Estimation." *Journal of Derivatives* 7, no. 3 (Spring 2000): 23–38.

Practice Questions and Problems (Answers at End of Book)

- 12.1 What assumption is being made when VaR is calculated using the historical simulation approach and 500 days of data?
- 12.2 Show that when λ approaches 1 the weighting scheme in Section 12.3.1 approaches the basic historical simulation approach.
- 12.3 Suppose we estimate the one-day 95% VaR from 1,000 observations (in millions of dollars) as 5. By fitting a standard distribution to the observations, the probability density function of the loss distribution at the 95% point is estimated to be 0.01. What is the standard error of the VaR estimate?
- 12.4 The one-day 99% VaR for the four-index example is calculated in Section 12.1 as \$422,291. Look at the underlying spreadsheets on the author's website and calculate (a) the 95% one-day VaR, (b) the 95% one-day ES, (c) the 97% one-day VaR, and (d) the 97% one-day ES.
- 12.5 Use the spreadsheets on the author's website to calculate a one-day 99% VaR and the one-day 99% ES using the basic methodology in Section 12.1, if the portfolio in Section 12.1 is equally divided between the four indices.
- 12.6 The weighting-of-observations procedure in Section 12.3.1 gives the one-day 99% VaR equal to \$653,541 and the one-day ES as \$833,228. Use the spreadsheets on the author's website to calculate VaR and ES when the λ parameter in this procedure is changed from 0.995 to 0.99.
- 12.7 The simplified volatility-scaling for portfolio procedure in Section 12.3.3 gives the one-day 99% VaR equal to \$517,195 and the one-day ES as \$650,354. Use the spreadsheets on the author's website to calculate VaR and ES when the λ parameter in this procedure is changed from 0.94 to 0.96.
- 12.8 In the application of extreme value theory (EVT) in Section 12.6, what is the probability that the loss will exceed \$400,000?
- 12.9 In the application of EVT in Section 12.6, what is the one-day VaR with a confidence level of 97%?
- 12.10 Change u from 160 to 150 in the application of EVT in Section 12.6. How does this change the maximum likelihood estimates of ξ and β ? How does it change the one-day 99% VaR and the one-day 99% ES when the confidence limit is (a) 99% and (b) 99.9%?
- 12.11 Carry out an extreme value theory analysis on the data from the volatility-scaling procedure in Table 12.7 and on the author's website. Use $u = 250$. What are the best fit values of ξ and β ? Calculate the one-day VaR and a one-day ES with a 99% and 99.9% confidence level. What is the probability of a loss greater than \$600,000?

Further Questions

- 12.12 Suppose that a one-day 97.5% VaR is estimated as \$13 million from 2,000 observations. The one-day changes are approximately normal with mean zero and standard deviation \$6 million. Estimate a 99% confidence interval for the VaR estimate.
- 12.13 Suppose that the portfolio considered in Section 12.1 has (in \$000s) 3,000 in S&P 500, 3,000 in FTSE, 1,000 in CAC 40, and 3,000 in Nikkei 225. Use the spreadsheet on the author's website to calculate what difference this makes to:
- (a) The one-day 99% VaR and ES that are calculated in Section 12.1
 - (b) The one-day 99% VaR and ES that are calculated using the weighting-of-observations procedure in Section 12.3.1 with $\lambda = 0.995$.
 - (c) The one-day 99% VaR and ES that are calculated using the two volatility-scaling procedures in Sections 12.3.2 and 12.3.3 with $\lambda = 0.94$ (assume that the initial variance when EWMA is applied is the sample variance).
 - (d) The one-day 99% VaR and ES that are calculated using extreme value theory and equal weightings in Section 12.6.
- 12.14 Investigate the effect of applying extreme value theory to the volatility-adjusted results in Section 12.3.2 with $u = 350$.
- 12.15 The weighting-of-observations procedure in Section 12.3.1 gives the one-day 99% VaR equal to \$653,541 and the one-day 99% ES as \$833,228. Use the spreadsheets on the author's website to calculate these measures when the λ parameter in this procedure is changed from 0.995 to 0.99.
- 12.16 The volatility-scaling procedure in Section 12.3.2 gives the one-day 99% VaR equal to \$566,492 and the 99% ES as \$705,855. Use the spreadsheets on the author's website to calculate VaR and ES when the λ parameter in this procedure is changed from 0.94 to 0.92.
- 12.17 Values for the NASDAQ Composite index during the 1,500 days preceding March 31, 2022, can be downloaded from the author's website. Calculate the one-day 99% VaR and one-day 99% ES on March 31, 2022, for a \$10 million portfolio invested in the index using:
- (a) The basic historical simulation approach
 - (b) The exponential weighting scheme in Section 12.3.1 with $\lambda = 0.995$
- Discuss the reasons for the differences between the results you get.

Chapter 13

Model-Building Approach

An alternative to the historical simulation approach for calculating risk measures such as value at risk (VaR) and expected shortfall (ES) is the *model-building approach*, sometimes also referred to as the *variance–covariance approach*. This involves assuming a model for the joint distribution of changes in market variables and using historical data to estimate the model parameters.

The model-building approach is ideally suited to the situation where the change in value of a portfolio is linearly dependent on changes in the values of the underlying market variables (stock prices, exchange rates, interest rates, etc.). In this case, if the daily changes in the underlying market variables are assumed to be multivariate normal, it is computationally much faster than historical simulation. It is in essence an extension of Harry Markowitz’s pioneering work in portfolio theory (see Section 1.1). The probability distribution of the change in the value of the portfolio is normal, and the mean and standard deviation of the change in the value of the portfolio can be calculated analytically from the mean and standard deviation of changes in the market variables and the correlations between those changes.

When a portfolio includes options and other non-linear derivatives, the probability distribution of the change in the value of a portfolio over a short period of time can no longer reasonably be assumed to be normal. We will discuss ways non-linearities can be handled when we discuss derivatives in Chapter 15. We will discuss how interest rates are handled by the model-building approach in Chapter 14.

13.1 The Basic Methodology

We start by considering how VaR is calculated using the model-building approach in a very simple situation where the portfolio under consideration consists of a position in a single stock. The portfolio we consider is one consisting of shares in Microsoft valued at \$10 million. We suppose that the time horizon is 10 days and the VaR confidence level is 99%, so we are interested in the loss level over 10 days that we are 99% confident will not be exceeded. Initially, we consider a one-day time horizon.

We assume that the volatility of Microsoft is 2% per day (corresponding to about 32% per year).¹ Because the size of the position is \$10 million, the standard deviation of daily changes in the value of the position is 2% of \$10 million, or \$200,000.

It is customary in the model-building approach to assume that the expected change in a market variable over the time period considered is zero. This is not exactly true, but it is a reasonable assumption. The expected change in the price of a market variable over a short time period is generally small when compared to the standard deviation of the change. Suppose, for example, that Microsoft has an expected return as high as 20% per annum. Over a one-day period, the expected return is $0.20/252$, or about 0.08%, whereas the standard deviation of the return is 2%. Over a 10-day period, the expected return is about $0.08\% \times 10$, or 0.8%, whereas the standard deviation of the return is $2\sqrt{10}$, or about 6.3%.

So far, we have established that the change in the value of the portfolio of Microsoft shares over a one-day period has a standard deviation of \$200,000 and (at least approximately) a mean of zero. We assume that the change is normally distributed.² Because $N(-2.326) = 0.01$, there is a 1% probability that a normally distributed variable will decrease in value by more than 2.326 standard deviations. Equivalently, we are 99% certain that a normally distributed variable will not decrease in value by more than 2.326 standard deviations. The one-day 99% VaR for our portfolio consisting of a \$10 million position in Microsoft is therefore

$$2.326 \times 200,000 = \$465,300$$

Assuming that the changes in Microsoft's stock price returns on successive days are independent, we can assume that the standard deviation of the return over 10 days is $0.02 \times \sqrt{10}$ and that the 10-day return is normally distributed. This gives the 10-day 99% VaR for Microsoft as

$$10,000,000 \times 0.02 \times \sqrt{10} \times 2.326 = \$1,471,300$$

¹As discussed in Section 8.1, in risk management calculations volatility is usually measured per day whereas in option pricing it is measured per year. A volatility per day can be converted to a volatility per year by multiplying by $\sqrt{252}$, or about 16.

²We could assume that the price of Microsoft is lognormal tomorrow. Because one day is such a short period of time, this is almost indistinguishable from the assumption we do make that the change in the stock price between today and tomorrow is normal.

The 10-day 99% ES is given by equation (11.2) with $\sigma = 200,000\sqrt{10} = 632,500$, $Y = 2.326$, and $X = 0.99$. It is \$1,686,000.

Consider next a portfolio consisting of a \$5 million position in AT&T, and suppose the daily volatility of AT&T is 1% (approximately 16% per year). A similar calculation to that for Microsoft shows that the standard deviation of the change in the value of the portfolio in one day is

$$5,000,000 \times 0.01 = 50,000$$

Assuming that the change is normally distributed, the one-day 99% VaR is

$$50,000 \times 2.326 = \$116,300$$

and the 10-day 99% VaR is

$$116,300 \times \sqrt{10} = \$367,800$$

The 10-day 99% ES is, from equation (11.2), \$421,400.

13.1.1 Two-Asset Case

Now consider a portfolio consisting of both \$10 million of Microsoft shares and \$5 million of AT&T shares. We suppose that the returns on the two shares have a correlation of 0.3. A standard result in statistics tells us that, if two variables X and Y have standard deviations equal to σ_X and σ_Y with the coefficient of correlation between them being equal to ρ , then the standard deviation of $X + Y$ is given by

$$\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y}$$

To apply this result, we set X equal to the change in the value of the position in Microsoft over a one-day period and Y equal to the change in the value of the position in AT&T over a one-day period, so that

$$\sigma_X = 200,000 \quad \sigma_Y = 50,000$$

The standard deviation of the one-day change in the value of the portfolio we are considering is therefore

$$\sqrt{200,000^2 + 50,000^2 + 2 \times 0.3 \times 200,000 \times 50,000} = 220,227$$

The mean change is assumed to be zero. If we further assume that the joint distribution of the returns from Microsoft and AT&T is bivariate normal, the change in the portfolio is normally distributed. The one-day 99% VaR is therefore

$$220,227 \times 2.326 = \$512,300$$

When a period of 10 days is considered, both σ_X and σ_Y increase by a multiplicative factor of $\sqrt{10}$. As a result, VaR also increases by a multiplicative factor of $\sqrt{10}$ to become \$1,620,100. The 10-day 99% ES is given by equation (11.2) with $\sigma = 220,227\sqrt{10}$, $Y = 2.326$, and $X = 0.99$. It is \$1,856,100.

13.1.2 The Benefits of Diversification

In the example we have just considered:

1. The 10-day 99% VaR for the portfolio of Microsoft shares is \$1,471,300.
2. The 10-day 99% VaR for the portfolio of AT&T shares is \$367,800.
3. The 10-day 99% VaR for the portfolio of both Microsoft and AT&T shares is \$1,620,100.

The amount

$$(\$1,471,300 + \$367,800) - \$1,620,100 = \$219,000$$

represents the benefits of diversification. If Microsoft and AT&T were perfectly correlated, the VaR for the portfolio of both Microsoft and AT&T would equal the VaR for the Microsoft portfolio plus the VaR for the AT&T portfolio. Less than perfect correlation leads to some of the risk being “diversified away.” This is true of ES as well as VaR.³

13.2 Generalization

The examples we have just considered are simple illustrations of the use of the linear model for calculating VaR. Suppose that we have a portfolio worth P that is dependent on n market variables. Following the market’s usual terminology, we will refer to market variables as *risk factors*. Examples of risk factors are equity prices, commodity prices, or exchange rates. (We defer consideration of interest rates until the next chapter.) We suppose that, to a good approximation, the change in the value of the portfolio is linearly related to proportional changes in the risk factors so that

$$\Delta P = \sum_{i=1}^n \delta_i \Delta x_i \quad (13.1)$$

where ΔP is the dollar change in the value of the whole portfolio in one day and Δx_i is the proportional change in the i th risk factor in one day.

The parameter, δ_i , is a variation on the delta risk measure that is used for derivatives and will be explained in Chapter 15. The delta of a position with respect to a risk factor is normally defined as the ratio $\Delta P / \Delta S$ where ΔS is a small change in the value of the risk factor (with all other risk factors remaining the same) and ΔP is the resultant change in the value of the portfolio. The parameter, δ_i , we use here is $\Delta P / \Delta x_i$ where Δx_i is a small *percentage change* in the value of the risk factor (again with all other risk factors remaining the same) and ΔP is the resultant dollar change in the value of the portfolio.

If we assume that the Δx_i in equation (13.1) are multivariate normal, ΔP is normally distributed. To calculate VaR, we therefore need to calculate only the mean and standard

³As discussed in Section 11.5, VaR does not always reflect the benefits of diversification. For non-normal distributions, the VaR of two portfolios considered jointly can be greater than the sum of their VaRs. ES does not have this disadvantage.

deviation of ΔP . We assume, as discussed in the previous section, that the expected value of each Δx_i is zero. This implies that the mean of ΔP is zero.

The way the standard deviation of ΔP is calculated is an extension of the two-asset example in the previous section. We define σ_i as the standard deviation of Δx_i , and ρ_{ij} is the coefficient of correlation between Δx_i and Δx_j . (Because we are considering one day, σ_i is the volatility per day.) The standard deviation of ΔP is given by

$$\sigma_P = \sqrt{\sum_{i=1}^n \sum_{j=1}^n \rho_{ij} \delta_i \delta_j \sigma_i \sigma_j} \quad (13.2)$$

This equation can also be written as

$$\sigma_P = \sqrt{\sum_{i=1}^n \delta_i^2 \sigma_i^2 + \sum_{i \neq j} \rho_{ij} \delta_i \delta_j \sigma_i \sigma_j}$$

or as

$$\sigma_P = \sqrt{\sum_{i=1}^n \delta_i^2 \sigma_i^2 + 2 \sum_{j < i} \rho_{ij} \delta_i \delta_j \sigma_i \sigma_j}$$

Yet another way of writing this key result is

$$\sigma_P^2 = \sum_{i=1}^n \sum_{j=1}^n \text{cov}_{ij} \delta_i \delta_j \quad (13.3)$$

where cov_{ij} is the covariance between Δx_i and Δx_j . Using matrix notation, this becomes

$$\sigma_P^2 = \boldsymbol{\delta}^T C \boldsymbol{\delta}$$

where $\boldsymbol{\delta}$ is the (column) vector whose i th element is δ_i , C is the variance-covariance matrix (see Section 9.3), and $\boldsymbol{\delta}^T$ is the transpose of $\boldsymbol{\delta}$.

The standard deviation of the change over T days is $\sigma_P \sqrt{T}$, and the VaR for a T -day time horizon and an $X\%$ confidence level is $N^{-1}(X) \sigma_P \sqrt{T}$. The ES for a T -day horizon and an $X\%$ confidence level is, from equation (11.2),

$$\sigma_P \sqrt{T} \frac{e^{-Y^2/2}}{\sqrt{2\pi}(1-X)}$$

where $Y = N^{-1}(X)$. In these results, N^{-1} is the inverse cumulative normal distribution function (NORMSINV in Excel).

Example 13.1

In the example considered in the previous section, the portfolio, P , had \$10 million invested in the first asset (Microsoft) and \$5 million invested in the second asset (AT&T). In this case, when amounts are measured in millions of dollars, $\delta_1 = 10$, $\delta_2 = 5$, and

$$\Delta P = 10\Delta x_1 + 5\Delta x_2$$

Also, $\sigma_1 = 0.02$, $\sigma_2 = 0.01$, and $\rho_{12} = 0.3$ so that

$$\sigma_p^2 = 10^2 \times 0.02^2 + 5^2 \times 0.01^2 + 2 \times 10 \times 5 \times 0.3 \times 0.02 \times 0.01 = 0.0485$$

and $\sigma_p = 0.220$. This is the standard deviation of the change in the portfolio value per day (in millions of dollars). The 10-day 99% VaR is $2.326 \times 0.220 \times \sqrt{10} = \1.62 million. The 10-day ES is:

$$0.220 \times \sqrt{10} \frac{e^{(-2.326^2/2)}}{\sqrt{2\pi} \times 0.01}$$

or \$1.86 million. These results agree with the calculations in the previous section.

13.2.1 Relation to Markowitz

The portfolio return in one day is $\Delta P/P$. From equation (13.2), the variance of this is

$$\sum_{i=1}^n \sum_{j=1}^n \rho_{ij} w_i w_j \sigma_i \sigma_j$$

where $w_i = \delta_i/P$. When the portfolio consists of long and short positions in n assets with the i th risk factor being the value of the i th asset, δ_i is the value of the investment in the i th asset (as shown in Example 13.1) and w_i is therefore the proportion of the portfolio invested in the i th asset. This version of equation (13.2) corresponds to the work of Markowitz and is often used by portfolio managers to relate the standard deviation of the return on a portfolio to the standard deviation of the returns on the components of the portfolio and the correlations between the returns on the components of the portfolio. (See Section 1.1.)

13.3 The Four-Index Example Revisited

As an example of the application of equation (13.2) or (13.3), return to the example considered in Section 12.1. This involved a portfolio on July 8, 2020, consisting of a \$4 million investment in the S&P 500, a \$3 million investment in the FTSE 100, a \$1 million investment in the CAC 40, and a \$2 million investment in the Nikkei 225. Daily returns were collected over 500 days ending on July 8, 2020. (Data and calculations presented here are on the author's website: www-2.rotman.utoronto.ca/~hull/RMFI/VaRExample.)

The correlation matrix, when calculations are carried out in the usual way with the same weight for all 500 returns, is shown in Table 13.1. The covariance matrix is shown in Table 13.2.

From equation (13.3), this variance-covariance matrix gives the variance of the portfolio losses (\$000s) as 14,406.193. The standard deviation is the square root of this, or 120.03. The one-day 99% VaR is therefore

$$2.326 \times 120.03 = 279.222$$

Table 13.1 Correlation Matrix on July 8, 2020, Calculated by Giving the Same Weight to Each of the Last 500 Daily Returns (variable 1 is S&P 500; variable 2 is FTSE 100; variable 3 is CAC 40; variable 4 is Nikkei 225)

$$\begin{pmatrix} 1 & 0.415 & 0.694 & 0.368 \\ 0.415 & 1 & 0.656 & 0.566 \\ 0.694 & 0.656 & 1 & 0.482 \\ 0.368 & 0.566 & 0.482 & 1 \end{pmatrix}$$

Table 13.2 Covariance Matrix on July 8, 2020, Calculated by Giving the Same Weight to Each of the Last 500 Daily Returns (variable 1 is S&P 500; variable 2 is FTSE 100; variable 3 is CAC 40; variable 4 is Nikkei 225)

$$\begin{pmatrix} 0.0002751 & 0.0000942 & 0.0001771 & 0.0000801 \\ 0.0000942 & 0.0001868 & 0.0001380 & 0.0001016 \\ 0.0001771 & 0.0001380 & 0.0002369 & 0.0000974 \\ 0.0000801 & 0.0001016 & 0.0000974 & 0.0001726 \end{pmatrix}$$

or \$279,222, and the one-day 99% ES is

$$120.03 \times \frac{e^{-2.326^2/2}}{\sqrt{2\pi} \times 0.01} = 319.894$$

or \$319,894. These values are much lower than the values given by the historical simulation approach in Chapter 12. This is because the latter is greatly affected by a handful of large losses occurring in March 2020.

13.3.1 Use of EWMA

Instead of calculating variances and covariances by giving equal weight to all observed returns, we now use the exponentially weighted moving average (EWMA) method with $\lambda = 0.94$. This gives the variance-covariance matrix in Table 13.3.⁴ From equation (13.3), the variance of portfolio losses (\$000s) is 16,903.752. The standard deviation is the square root of this, or 130.014. The one-day 99% VaR is therefore

$$2.326 \times 130.014 = 302.459$$

or \$302,459, and the one-day ES is, from equation (11.2), 346,516. Table 13.4 shows the estimated volatilities of the indices. Volatilities are higher when EWMA is used because

⁴In the EWMA calculations, the variance is initially set equal to the population variance. But all reasonable starting variances give similar results because in this case all we are interested in is the final variance.

Table 13.3 Covariance Matrix on July 8, 2020, When EWMA with $\lambda = 0.94$ Is Used (variable 1 is S&P 500; variable 2 is FTSE 100; variable 3 is CAC 40; variable 4 is Nikkei 225)

$$\begin{pmatrix} 0.0003089 & 0.0000681 & 0.0002160 & 0.0000678 \\ 0.0000681 & 0.0002729 & 0.0001878 & 0.0001586 \\ 0.0002160 & 0.0001878 & 0.0003584 & 0.0001464 \\ 0.0000678 & 0.0001586 & 0.0001464 & 0.0002707 \end{pmatrix}$$

Table 13.4 Volatilities (% per day) on July 8, 2020, for the Equal Weighting and EWMA Approaches

	S&P 500	FTSE 100	CAC 40	Nikkei 225
Equal weighting	1.66	1.37	1.54	1.31
EWMA	1.76	1.65	1.89	1.65

Table 13.5 Correlation Matrix on July 8, 2020, When EWMA Method Is Used (variable 1 is S&P 500; variable 2 is FTSE 100; variable 3 is CAC 40; variable 4 is Nikkei 225)

$$\begin{pmatrix} 1 & 0.235 & 0.649 & 0.235 \\ 0.235 & 1 & 0.601 & 0.584 \\ 0.649 & 0.601 & 1 & 0.470 \\ 0.235 & 0.584 & 0.470 & 1 \end{pmatrix}$$

more weight is given to the March 2020 period. Comparing Table 13.5 with Table 13.1, we see that correlations are also lower when EWMA is used. Normally correlations increase in stressed market conditions. In this case, it appears that movements in indices were less closely aligned during the first half of 2020.

13.4 Extensions of the Basic Procedure

In this section we continue to consider the situation where a portfolio is linearly dependent on the underlying risk factors and explain some extensions of the model in equations (13.2).

13.4.1 Stressed Measures

Suppose we are interested in stressed VaR or stressed ES. As discussed in Section 12.1.3, this is an estimate based on data from a stressed period in the past, not from a recent

period. The methods we have proposed so far in this chapter can be modified by estimating volatilities and correlations using data from a stressed period rather than from a recent period.

13.4.2 Non-Normal Distributions

In Section 11.9 we explained that an approximate formula for aggregating the VaRs for n portfolios is

$$\text{VaR}_{\text{total}} = \sqrt{\sum_{i=1}^n \sum_{j=1}^n \rho_{ij} \text{VaR}_i \text{VaR}_j} \quad (13.4)$$

where VaR_i is the VaR for the i th portfolio, $\text{VaR}_{\text{total}}$ is the total VaR, and ρ_{ij} is the correlation between losses from the i th and j th portfolios. The formula shows that the method for aggregating standard deviations can also, as an approximation, be used to aggregate percentiles.

The result suggests a way the model-building approach can be extended to allow for return distributions being non-normal. Suppose that, based on historical data, we estimate that the one-percentile point of the 10-day return distribution for Microsoft is -17% so that the 10-day 99% VaR of \$10 million invested in Microsoft is \$1,700,000 (rather than the \$1,471,300 calculated assuming normally distributed returns in Section 13.1). Suppose further that we estimate that the one-percentile point of the 10-day return distribution for AT&T is -10% so that the 10-day 99% VaR for \$5 million invested in AT&T is \$500,000 (rather than the \$367,800 estimated assuming normally distributed returns). Using equation (13.7) we would adjust our estimate of the VaR of the portfolio of Microsoft and AT&T to

$$\sqrt{1,700,000^2 + 500,000^2 + 2 \times 0.3 \times 1,700,000 \times 500,000} = 1,910,500$$

13.5 Risk Weights and Weighted Sensitivities

When equations such as (13.2) and (13.4) are used to calculate VaR or ES, the standard deviation of the daily change in the portfolio is multiplied by a constant. For example, when the 10-day 99% VaR is calculated, it is multiplied by $N^{-1}(0.99) \times \sqrt{10} = 7.36$. When the 20-day 99% ES is calculated, it is multiplied by

$$\frac{\sqrt{20} \exp\{-[N^{-1}(0.99)]^2/2\}}{\sqrt{2\pi} \times 0.01} = 11.92$$

Suppose that β is the multiplier so that:

$$\text{Risk Measure} = \beta \sqrt{\sum_{i=1}^n \sum_{j=1}^n \delta_i \delta_j \rho_{ij} \sigma_i \sigma_j}$$

Defining $W_i = \beta\sigma_i$. This can be written:

$$\text{Risk Measure} = \sqrt{\sum_{i=1}^n \sum_{j=1}^n \delta_i \delta_j \rho_{ij} W_i W_j} \quad (13.5)$$

The W -parameters are referred to as *risk weights*, and the δ -parameters are referred to as *weighted sensitivities*.

As we shall see in later chapters, this formulation is used for

1. Determining capital for market risk in the Fundamental Review of the Trading Book using the standardized approach
2. Determining initial margin in bilaterally cleared transactions

13.6 Non-Linearity

It should be noted that the results in this chapter do not apply when the relationship between the portfolio value and the risk factors is non-linear. This is because the change in the portfolio value cannot be assumed to be normally distributed. From Appendix G, for a portfolio dependent on a single risk factor, x , an approximation is

$$\Delta P = \delta \Delta x + \frac{1}{2} \gamma \Delta x^2 \quad (13.6)$$

where δ and γ are the first and second partial derivatives of the portfolio value, P , with respect to x . When there are n risk factors, x_i ($1 \leq i \leq n$), an approximation is

$$\Delta P = \sum_{i=1}^n \delta_i \Delta x_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \gamma_{ij} \Delta x_i \Delta x_j \quad (13.7)$$

where

$$\delta_i = \frac{\partial P}{\partial x_i}$$

$$\gamma_{ij} = \frac{\partial^2 P}{\partial x_i \partial x_j}$$

13.7 Model-Building vs. Historical Simulation

In the preceding chapter and in this one, we have discussed two methods for estimating VaR: the historical simulation approach and the model-building approach. The advantages of the model-building approach are that results can be produced very quickly, and they can easily be used in conjunction with volatility and correlation updating procedures such as those described in Chapters 8 and 9. (As mentioned in Section 12.3, volatility changes can be incorporated into the historical simulation approach—but in a

rather more artificial way.) The main disadvantage of the model-building approach, as we have described it in this Sections 13.1 to 13.5, is that the change in the value of the portfolio is linearly related to the proportional or actual changes in the risk factors and the daily changes in the risk factors are assumed to have a multivariate normal distribution. In practice, daily changes in risk factors often have distributions that are quite different from normal. (See, for example, Table 8.1.) A user of the model-building approach can then use the approach in Section 13.4.2, hoping that some form of the central limit theorem of statistics applies so that the probability distribution of daily gains/losses on a large portfolio is normally distributed, even though the gains/losses on the component parts of the portfolio are not normally distributed. In Chapter 15 we will explain some extension of the model-building approach that can be used to handle non-linearities.

Summary

Whereas historical simulation lets the data determine the joint probability distribution of daily percentage changes in risk factors, the model-building approach assumes a particular form for this distribution. The most common assumption is that percentage changes or actual changes in the risk factors have a multivariate normal distribution. For situations where the change in the value of the portfolio is linearly dependent on changes in the risk factors, VaR and ES can then be calculated very quickly.

The model-building approach is frequently used for investment portfolios where there are no derivatives. As we will see in Chapter 15, it is much more difficult to use it when a portfolio includes derivatives.

Further Reading

Frye, J. “Principles of Risk: Finding VAR Through Factor-Based Interest Rate Scenarios.” In *VAR: Understanding and Applying Value-at-Risk*, edited by Sue Grayling. London: Risk Publications, 1997: 275–288.

Hull, J. C., and A. White. “Value at Risk When Daily Changes in Market Variables Are Not Normally Distributed.” *Journal of Derivatives* 5 (Spring 1998): 9–19.

Jamshidian, F., and Y. Zhu. “Scenario Simulation Model: Theory and Methodology.” *Finance and Stochastics* 1 (1997): 43–67.

Practice Questions and Problems (Answers at End of Book)

- 13.1 Consider a position consisting of a \$100,000 investment in asset A and a \$100,000 investment in asset B. Assume that the daily volatilities of both assets are 1% and that the coefficient of correlation between their returns is 0.3. What are the five-day 97% VaR and ES for the portfolio?

- 13.2 A financial institution owns a portfolio of instruments dependent on the U.S. dollar–sterling exchange rate. The delta of the portfolio with respect to percentage changes in the exchange rate in equation (13.1) is 3.9. If the daily volatility of the exchange rate is 0.7% and the linear model in equation (13.1) is assumed, estimate the 10-day 99% VaR.
- 13.3 Suppose that in Problem 13.2 the relationship is non-linear and the second partial derivative of the portfolio value with respect to the exchange rate is 4.3. Use a Taylor series to derive an approximate relationship between the change in the portfolio value and the percentage change in the exchange rate?
- 13.4 Explain how the terms *risk weights* and *risk sensitivity* are used in connection with the model-building approach.
- 13.5 Suppose that the daily change in the value of a portfolio is, to a good approximation, linearly dependent on two factors. The delta of a portfolio in equation (13.1) with respect to the first factor is 6 and the delta with respect to the second factor is -4 . The standard deviations of the factor are 20 and 8, respectively. What is the five-day 90% VaR?
- 13.6 Explain why the linear model can provide only approximate estimates of VaR for a portfolio containing options.
- 13.7 Some time ago, a company entered into a forward contract to buy £1 million for \$1.5 million. The contract now has six months to maturity. The daily volatility of a six-month zero-coupon sterling bond (when its price is translated to dollars) is 0.06%, and the daily volatility of a six-month zero-coupon dollar bond is 0.05%. The correlation between returns from the two bonds is 0.8. The current exchange rate is 1.53. Calculate the standard deviation of the change in the dollar value of the forward contract in one day. What is the 10-day 99% VaR? Assume that the six-month interest rate in both sterling and dollars is 5% per annum with continuous compounding.
- 13.8 The calculations in Section 13.3 assume that the investments in the S&P 500, FTSE 100, CAC 40, and Nikkei 225 are \$4 million, \$3 million, \$1 million, and \$2 million, respectively. How do the VaR and ES change if the investment is \$2.5 million in each index? Carry out calculations when (a) volatilities and correlations are estimated using the equally weighted model and (b) when they are estimated using the EWMA model with $\lambda = 0.94$. Use the spreadsheets on the author's website.
- 13.9 What is the effect of changing λ from 0.94 to 0.97 in the EWMA calculations in Section 13.3? Use the spreadsheets on the author's website.

Further Questions

- 13.10 Consider a position consisting of a \$300,000 investment in gold and a \$500,000 investment in silver. Suppose that the daily volatilities of these two assets are 1.8% and 1.2%, respectively, and that the coefficient of correlation between their returns

- is 0.6. What are the 10-day 97.5% VaR and ES for the portfolio? By how much does diversification reduce the VaR and ES?
- 13.11 Consider a portfolio of options on a single asset. Suppose that the delta of the portfolio (calculated with respect to actual changes) is 12, the value of the asset is \$10, and the daily volatility of the asset is 2%. What is the delta with respect to proportional changes? Estimate the one-day 95% VaR for the portfolio from the delta.
- 13.12 The calculations in Section 13.3 assume that the investments in the S&P 500, FTSE 100, CAC 40, and Nikkei 225 are \$4 million, \$3 million, \$1 million, and \$2 million, respectively. How do the VaR and ES change if the investments are \$3 million, \$3 million, \$1 million, and \$3 million, respectively? Carry out calculations when (a) volatilities and correlations are estimated using the equally weighted model and (b) when they are estimated using the EWMA model. What is the effect of changing λ from 0.94 to 0.90 in the EWMA calculations? Use the spreadsheets on the author's website.

Chapter 14

Interest Rate Risk

Interest rate risk is more difficult to manage than the risk arising from market variables such as equity prices, exchange rates, and commodity prices. One complication is that there are many different interest rates in any given currency (Treasury rates, interbank borrowing and lending rates, swap rates, and so on). Although these tend to move together, they are not perfectly correlated. Another complication is that we need more than a single number to describe the interest rate environment. We need a function describing the variation of the interest rate with maturity. This is known as the *term structure of interest rates* or the *yield curve*. The interest rate considered is usually a *zero-coupon interest rate*, sometimes just referred to as a “zero.” It is the rate of interest that would apply for the maturity being considered if all interest and principal are paid at the end (i.e., there are no periodic payments such as those that are usually made on bonds).

Consider, for example, the situation of a U.S. government bond trader. The trader’s portfolio is likely to consist of many bonds with different maturities. There is an exposure to movements in the one-year rate, the two-year rate, the three-year rate, and so on. The trader’s delta exposure is therefore more complicated than that of a gold trader, for example. He or she must be concerned with all the different ways in which the U.S. Treasury yield curve can change its shape through time.

This chapter starts by explaining some of the interest rates that are important to financial institutions. It then moves on to consider how the methods for calculating VaR and ES in Chapters 12 and 13 can be used for interest rate exposures. The management of net interest income is then considered. Finally, the chapter explains the use of duration and convexity measures.

This chapter does not cover issues such as the compounding frequency with which interest rates are measured and how the term structure of zero-coupon interest rates is calculated. These topics are covered in Appendices A and B.

14.1 Types of Rates

An interest rate in a particular situation defines the amount of money a borrower promises to pay the lender. For any given currency, many different types of interest rates are regularly quoted. These include mortgage rates, deposit rates, prime borrowing rates, and so on. One important factor influencing interest rates is credit risk. This is the risk that there will be a default by the borrower of funds, so that the interest and principal are not paid to the lender as promised. The higher the credit risk, the higher the interest rate that is promised by the borrower. The extra amount added to a risk-free interest rate to allow for credit risk is known as a *credit spread*. Interest rates are often expressed in basis points. One basis point is 0.01% per annum.

14.1.1 Treasury Rates

Treasury rates are the rates an investor earns on Treasury bills and Treasury bonds. These are the instruments used by a government to borrow in its own currency. Japanese Treasury rates are the rates at which the Japanese government borrows in yen; U.S. Treasury rates are the rates at which the U.S. government borrows in U.S. dollars; and so on. It is usually assumed that there is no chance that the government of a developed country will default on an obligation denominated in its own currency. A developed country's Treasury rates are therefore regarded as risk-free in the sense that an investor who buys a Treasury bill or Treasury bond is certain that interest and principal payments will be made as promised.

14.1.2 Overnight Rates

There is a great deal of borrowing and lending between banks in the overnight market and the rates paid are monitored. For the U.S. dollar, the most important overnight rate is the secured overnight funding rate (SOFR). As its name implies, this is a secured rate (i.e., backed by collateral). A similar secured overnight rate for the Swiss franc is the Swiss average rate overnight (SARON). Important unsecured overnight rates in the United Kingdom, the Eurozone, and Japan are the sterling overnight index average (SONIA), the euro short-term rate (ESTER), and the Tokyo overnight average rate (TONAR), respectively.

14.1.3 Reference Rates

Reference interest rates are important in financial markets. The parties to transactions frequently enter into contracts where the future interest rate paid or received is uncertain,

but will be calculated based on the value of an agreed reference interest rate. LIBOR used to be the most important reference rate. But as explained in Section 5.3.3, its use has now been discontinued. Regulators have suggested that reference rates be replaced by the overnight rates just mentioned.

Longer rates such as three-month rates, six-month rates, or one-year rates can be determined from overnight rates by compounding them daily. Suppose that the annualized SOFR rate for the i th business day in a period is r_i and the rate applies to d_i days. The annualized SOFR rate is a rate per 360 days. The annualized interest rate for the period is therefore

$$[(1 + r_1 \hat{d}_1)(1 + r_2 \hat{d}_2) \cdots (1 + r_n \hat{d}_n) - 1] \times \frac{360}{D}$$

where $\hat{d}_i = d_i/360$ and $D = \sum_i d_i$ is the number of days in the period. On most days $d_i = 1$, but weekends and holidays lead to overnight rates being applied to more than one day. (For example, on a Friday d_i will usually be 3.)

As mentioned in Section 5.3.3, there are important differences between the new reference rates proposed by regulators and the old ones. LIBOR rates are forward looking. They are determined at the beginning of the period to which they will apply. The new reference rates are backward looking in the sense that the rate for a period is not known until the end of the period when all the relevant overnight rates have been observed. Another difference is that the new rates are regarded as risk-free because they are derived from one-day loans to creditworthy financial institutions. LIBOR, by contrast, incorporates a credit spread. The spread between three-month LIBOR and a three-month rate based on overnight rates is usually about 10 basis points (0.1%), but it can be much higher in stressed market conditions. For example, it spiked to an all-time high of 364 basis points (3.64%) in the United States in October 2008 during the Global Financial Crisis.

If a bank offers a loan at a reference rate plus $x\%$, where x is a constant, it would usually like the rate to reflect ups and downs in average credit spreads. LIBOR, when used as a reference rate, did this, but the overnight rates (because they are essentially risk-free) do not. As mentioned in Section 5.3.3, this has led to the development of some alternatives to the rates proposed by regulators. One is the Bloomberg Short-Term Yield Index (BSBY), which estimates the average yield at which large banks in the United States access unsecured funds. Another is AMERIBOR, which reflects the actual borrowing costs of thousands of community and regional banks in the United States. It will be interesting to see which of the reference rates that have been proposed as replacements for LIBOR become most actively used.

14.1.4 Repo Rates

In a repo (or repurchase agreement), a financial institution that owns securities agrees to sell the securities for a certain price and to buy them back at a later time for a slightly higher price. The financial institution is obtaining a loan, and the interest it pays is the difference between the price at which the securities are sold and the price at which they are repurchased. The interest rate is referred to as the repo rate.

If structured carefully, a repo involves very little credit risk. If the borrower does not honor the agreement, the lending company simply keeps the securities. If the lending company does not keep to its side of the agreement, the original owner of the securities keeps the cash provided by the lending company. The most common type of repo is an overnight repo, which may be rolled over day to day. However, longer-term arrangements, known as term repos, are sometimes used.

14.2 Calculating Risk Measures

Term structures create complications for the application of the approaches for calculating VaR and ES in Chapters 12 and 13. One instrument in a company's portfolio might lead to a cash flow in 3.32 years so that the change in the value of the portfolio over the next short period of time depends on what will happen to the 3.32-year interest rate. Another instrument might lead to a cash flow at the 4.48-year point so that the financial institution also has an exposure to that rate. The 3.32-year rate and the 4.48-year rate will tend to move together, but they are not perfectly correlated.

It is clearly not possible to consider separately every single maturity to which a financial institution is exposed. Typically, the financial institution will calculate, from the yields on available instruments, zero-coupon interest rates for certain standard maturities such as three months, six months, one year, etc. The way this is done is outlined in Appendix B. It will monitor the movements in these interest rates and calculate other rates as needed using linear interpolation. It will do this for all the different term structures of interest rates to which it is exposed.

When the historical simulation approach is used, the rate for each standard maturity for each term structure is considered as a separate risk factor. The change in each of these rates in the i th scenario is the observed change between Day $i-1$ and Day i of the historical data. As discussed in Section 13.1 calculations are based on actual changes rather than percentage changes in risk factors. Required rates for nonstandard maturities can be calculated using linear interpolation between the rates for standard maturities as indicated in Appendix B.

When the model-building approach is used, the rate for each standard maturity on each term structure can also be defined as a separate risk factor. A delta for the rate at the i th point is calculated as $\Delta P / \Delta r$ where Δr is a small change in the rate and ΔP the change in the portfolio value when that rate is changed and others are kept fixed. Figure 14.1 considers a term structure defined by points corresponding to maturities 3 months, 6 months, 1 year, 2 years, 3 years, 5 years, 10 years, 15 years, 20 years, and 30 years, and shows the effect on the term structure making a small change to the five-year rate. The deltas for other points on the term structure are calculated similarly. All rates less than the shortest rate (three months in our example) are changed by the same amount Δr and rates between that rate and the next rate are changed in the same way as the rates between five years and ten years in Figure 14.1. Similarly, for the longest maturity rate, all rates greater than the longest rate are changed by Δr and the rates

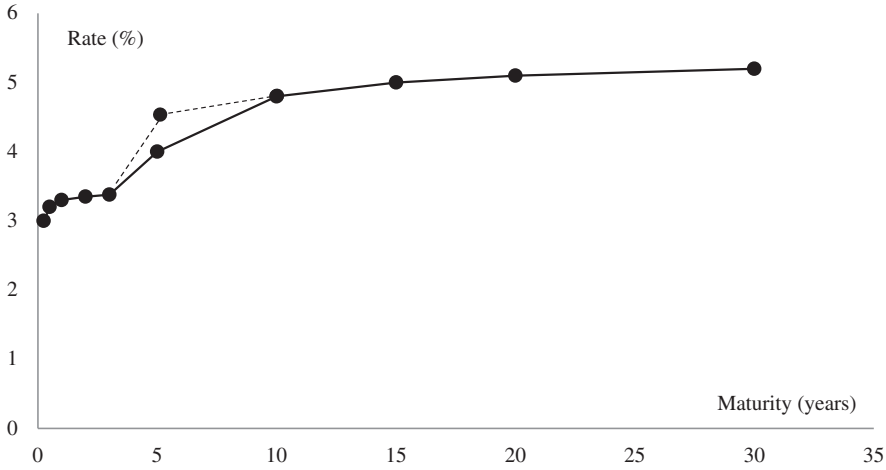


Figure 14.1 Five-Year Rate Is Changed with Other Rates Kept Constant

between the second-longest and the longest maturity are changed in the same way as rates between three years and five years in Figure 14.1. (This is consistent with the usual convention for constructing zero-coupon term structures described in Appendix B.) The result of all of this is that the sum on the deltas for the points on the term structure is the delta for a parallel shift in the whole term structure. The DV01 of a portfolio is defined as the effect on it of a one-basis-point parallel shift. The deltas we have described are in effect a way of dividing this into components, one for each point on the term structure.

Consider a portfolio whose value depends on only one term structure. Define δ_i as the delta for the i th point on the term structure. If σ_i is the standard deviation of a one-day change in the rate corresponding to this point and ρ_{ij} is the correlation between movements in rates at the i th and j th points, the standard deviation of the change in the value of the portfolio in one day is

$$\sigma_P = \sqrt{\sum_{i=1}^n \sum_{j=1}^n \rho_{ij} \delta_i \delta_j \sigma_i \sigma_j} \quad (14.1)$$

where n is the number of points on the term structure.

Any given future cash flow has a delta with respect to the two adjacent maturities. For example, a cash flow in 3.5 years has a delta with respect to the three-year rate and the five-year rate, but none with respect to the other rates. When the exposure to the 3.5-year rate is X , the nature of the changes indicated in Figure 14.1 shows that this translates to an exposure of $0.75X$ to the three-year rate and $0.25X$ to the five-year rate.

Note that equation (14.1) is the same as equation (13.2). The only difference is that equation (13.2) works with proportional changes in the risk factors (the δ -variables measure the impact on the portfolio of a proportional changes, and the σ -variables are volatilities), whereas equation (14.1) works with actual changes in the risk factors (the δ -variables measure the impact of actual changes in the risk factors and the σ -variables

are standard deviations). Equations (13.2) and (14.1) can be extended to cover a portfolio where proportional changes are considered for some risk factors and actual changes are considered for others.

Suppose now that portfolio P depends on K term structures. Define δ_{ik} , σ_{ik} , and ρ_{ijk} as the values of δ_i , σ_i , and ρ_{ij} for term structure k ($1 \leq k \leq K$). Also define

$$V_k^2 = \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} \rho_{ijk} \delta_{ik} \delta_{jk} \sigma_{ik} \sigma_{jk}$$

$$U_k = \sum_{i=1}^{N_k} \delta_{ik} \sigma_{ik}$$

where N_k is the number of vertices used for term structure k .

It is usual to describe the correlation between the two term structures with a single parameter. Suppose that $\rho(k_1, k_2)$ is the correlation between term structures k_1 and k_2 . Two ways of defining it are as follows:

1. The correlation between rate i for term structure k_1 and rate j for term structure k_2 is $\rho(k_1, k_2)$ for all i and j .
2. The change in the value of the portfolio due to movements in term structure k_1 and the change in the value of the portfolio due to movements in term structure k_2 have a correlation of $\rho(k_1, k_2)$.

The first definition leads to

$$\sigma_P = \sqrt{\sum_k V_k^2 + \sum_{k_1 \neq k_2} \rho(k_1, k_2) U_{k_1} U_{k_2}} \quad (14.2)$$

whereas the second gives

$$\sigma_P = \sqrt{\sum_k V_k^2 + \sum_{k_1 \neq k_2} \rho(k_1, k_2) V_{k_1} V_{k_2}} \quad (14.3)$$

Example 14.1

Suppose that a portfolio has exposures to two different interest rate term structures. The standard deviations of rate moves per day and the delta exposures to rate moves are shown in Table 14.1. For example, the standard deviation of daily changes in the two-year rate in term structure 1 is 5.6 basis points (0.056%), while that for term structure 2 is 11.4 basis points. The impact of a one-basis-point change in the two-year rate is \$85 million for term structure 1 and \$65 million for term structure 2. We assume that correlations ρ_{ij} are the same for both term structures and are those shown in Table 14.2. Thus the correlation between the change in the one-year rate in term structure 1 and the change in the two-year rate in term structure 1 is 0.92. The same is true for term structure 2.

Table 14.1 Standard Deviations (in basis points) and Delta Exposures (in millions of dollars per basis point)

	3m	6m	1yr	2yr	3yr	5yr	10yr	15yr	20yr	30yr
TS1 SD per day	8.8	7.4	6.7	5.6	5.4	5.4	5.2	5.2	5.5	6.4
TS2 SD per day	10.2	10.8	12.0	11.4	11.0	11.4	10.0	11.2	11.2	11.3
Delta TS1	55	65	80	85	90	70	65	40	20	5
Delta TS2	85	75	70	65	50	45	40	30	20	20

Table 14.2 Correlations between Rates of Different Maturities

	3m	6m	1yr	2yr	3yr	5yr	10yr	15yr	20yr	30yr
3m	1.00	0.78	0.62	0.50	0.44	0.36	0.27	0.20	0.17	0.13
6m	0.78	1.00	0.84	0.74	0.67	0.57	0.44	0.37	0.35	0.30
1yr	0.62	0.84	1.00	0.92	0.86	0.76	0.63	0.55	0.53	0.47
2yr	0.50	0.74	0.92	1.00	0.98	0.89	0.75	0.69	0.66	0.60
3yr	0.44	0.67	0.86	0.98	1.00	0.96	0.83	0.78	0.75	0.69
5yr	0.36	0.57	0.76	0.89	0.96	1.00	0.92	0.89	0.86	0.81
10yr	0.27	0.44	0.63	0.75	0.83	0.92	1.00	0.98	0.96	0.93
15yr	0.20	0.37	0.55	0.69	0.78	0.89	0.98	1.00	0.99	0.97
20yr	0.17	0.35	0.53	0.66	0.75	0.86	0.96	0.99	1.00	0.99
30yr	0.13	0.30	0.47	0.60	0.69	0.81	0.93	0.97	0.99	1.00

Finally, we assume that the correlation parameter between the two term structures is 0.4. In millions of dollars, the results are:

$$U_1 = 3,529 \quad U_2 = 5,501$$

$$V_1 = 3,004.9 \quad V_2 = 4,604.3$$

The portfolio standard deviation per day given by equation (14.2) is 6,163.9, and the portfolio standard deviation given by equation (14.3) is 5,980.2. These must be multiplied by $\sqrt{10}$ and 2.326 to get the 10-day 99% VaR. The expected shortfall can be calculated using equation (11.2).

The calculations are shown on the author's website:

www-2.rotman.utoronto.ca/~hull/riskman.

14.3 Principal Components Analysis

A statistical tool for handling groups of highly correlated risk factors is principal components analysis (PCA). The approach is best illustrated with an example. We consider daily data on changes in 1-year, 2-year, 3-year, 5-year, 7-year, 10-year, 20-year, and 30-year U.S. Treasury rates between January 2010 and July 2020. Tables 14.3 and 14.4 show results produced by a PCA analysis. The columns labeled PC1, PC2, etc. show the eight

Table 14.3 Factor Loadings for U.S. Treasury Rates

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
1yr	0.083	-0.242	0.685	-0.682	-0.006	-0.025	-0.021	-0.004
2yr	0.210	-0.465	0.376	0.574	-0.517	-0.031	0.011	-0.008
3yr	0.286	-0.467	0.006	0.185	0.728	0.347	0.106	-0.074
5yr	0.386	-0.315	-0.332	-0.145	0.061	-0.604	-0.348	0.361
7yr	0.430	-0.099	-0.349	-0.265	-0.266	-0.008	0.263	-0.688
10yr	0.428	0.119	-0.153	-0.172	-0.269	0.515	0.254	0.589
20yr	0.426	0.394	0.172	0.099	0.027	0.244	-0.722	-0.205
30yr	0.411	0.478	0.323	0.204	0.234	-0.434	0.461	0.036

Table 14.4 Standard Deviation of Factor Scores

PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
11.54	3.55	1.78	1.25	0.91	0.69	0.62	0.57

components or factors describing the rate moves. The first factor, labeled PC1, represents a shift where all rates move in the same direction. When we have one unit of that factor, the 1-year rate increases by 0.083 basis points, the 2-year rate increases by 0.210 basis points, and so on. The second factor is shown in the column labeled PC2. It represents a “twist” or change in the slope of the yield curve. Rates with maturities between 1 year and 7 years move in one direction; rates with maturities between 10 and 30 years move in the other direction. The third factor corresponds to a “bowing” of the yield curve. Relatively short rates and relatively long rates move in one direction; the intermediate rates move in the other direction. The interest rate movement for a particular factor is known as factor loading. In our example, the first factor’s loading for the 1-year rate is 0.083.¹ (Note that the signs for factor loadings are somewhat arbitrary. We can change the signs for all factor loadings corresponding to a particular factor without changing the model. For example, if we did so for the first factor, -1 unit of that factor would result in the same yield curve changes as $+1$ unit does in Table 14.3.)

Because there are eight rates and eight factors, the interest rate changes observed on any given day can always be expressed as a linear sum of the factors by solving a set of eight simultaneous equations. The quantity of a particular factor in the interest rate changes on a particular day is known as the factor score for that day.

The importance of a factor is measured by the standard deviation of its factor score. The standard deviations of the factor scores in our example are shown in Table 14.4, and the factors are listed in order of their importance. The numbers in Table 14.4 are measured in basis points. A quantity of the first factor equal to 1 standard deviation, therefore, corresponds to the 1-year rate moving by $0.083 \times 11.54 = 0.96$ basis points; the 2-year rate moving by $0.210 \times 11.54 = 2.42$ basis points, and so on.

¹Factor loadings have the property that the sum of their squares for each factor is 1.0.

Software for carrying out the calculations underlying Tables 14.3 and 14.4 is on the author's website. The calculations are explained in Appendices H and I at the end of the book. The factors have the property that the factor scores are uncorrelated across the data. For instance, in our example, the first factor score is uncorrelated with the second factor score for the days considered. The variances of the factor scores have the property that they add up to the total variance of the data. From Table 14.4, the total variance of the original data (that is, sum of the variance of the observations on the 1-year rate, the variance of the observations on the 2-year rate, and so on) is

$$11.54^2 + 3.55^2 + 1.78^2 + \cdots + 0.57^2 = 152.5$$

From this it can be seen that the first factor accounts for $11.54^2/152.5 = 87.3\%$ of the variance in the original data; the first two factors account for

$$\frac{11.54^2 + 3.55^2}{152.5} = 95.6\%$$

of the variance in the data; the third factor accounts for a further 2.1% of the variance. This shows that most of the risk in interest rate moves is accounted for by the first two or three factors. It suggests that we can relate the risks in a portfolio of interest rate dependent instruments to movements in these factors instead of considering all eight interest rates.

The three most important factors from Table 14.3 are plotted in Figure 14.2.²

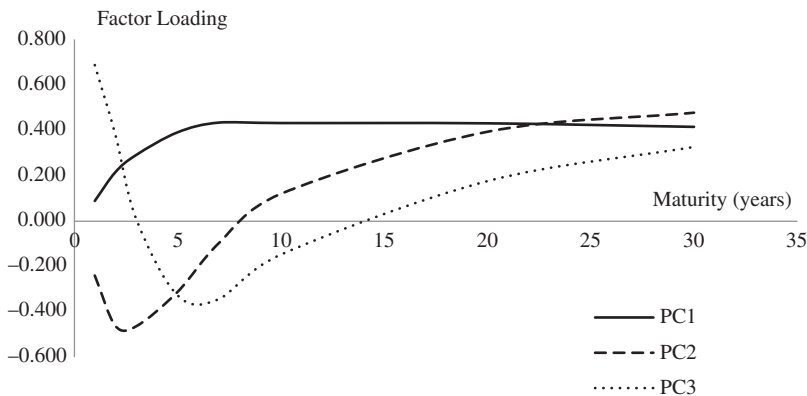


Figure 14.2 The Three Most Important Factors Driving Treasury Rates in the United States Using Daily Data between 2010 and 2020

²Results similar to those here, concerning the nature of the factors and the amount of the total risk they account for, are obtained when a principal components analysis is used to explain movement in almost any yield curve in any country.

Table 14.5 Change in Portfolio Value for a 1-Basis-Point Rate Move

2-Year Rate	3-Year Rate	5-Year Rate	7-Year Rate	10-Year Rate
+10	+4	−8	−7	+2

14.3.1 Calculating VaR and ES

To illustrate how a principal components analysis can be used to calculate VaR and ES, consider a portfolio with the exposures to interest rate moves shown in Table 14.5. A one-basis-point change in the 2-year rate causes the portfolio value to increase by \$10 million; a one-basis-point change in the 3-year rate causes it to increase by \$4 million, and so on. Suppose the first two factors are used to model rate moves. (As mentioned above, this captures 95.6% of the variance in rate moves.) Using the data in Table 14.2, the exposure to the first factor (measured in millions of dollars per factor score basis point) is

$$10 \times 0.210 + 4 \times 0.286 - 8 \times 0.386 - 7 \times 0.430 + 2 \times 0.428 = -1.99$$

and the exposure to the second factor is

$$10 \times (-0.465) + 4 \times (-0.467) - 8 \times (-0.315) - 7 \times (-0.099) + 2 \times 0.119 = -3.06$$

Suppose that f_1 and f_2 are the factor scores (measured in basis points). The change in the portfolio value is, to a good approximation, given by

$$\Delta P = -1.99f_1 - 3.06f_2$$

The factor scores are uncorrelated and have the standard deviations given in Table 14.2.

The standard deviation of ΔP is therefore

$$\sqrt{1.99^2 \times 11.54^2 + 3.06^2 \times 3.55^2} = 25.45$$

Assuming normally distributed factors, the one-day 99% VaR is $25.45 \times 2.326 = 59.2$. The one-day 99% ES is from equation 11.2:

$$25.45 \times \frac{e^{-2.326^2/2}}{\sqrt{2\pi} \times 0.01} = 67.9:$$

A principal components analysis can in theory be used for market variables other than interest rates. Suppose that a financial institution has exposures to a number of different stock indices. A principal components analysis can be used to identify factors describing movements in the indices, and the most important of these can be used to replace the market indices in a VaR analysis. How effective a principal components analysis is for a group of market variables depends on how closely correlated they are.

14.4 The Management of Net Interest Income

A key risk management activity for a bank is the management of net interest income. As explained in Section 2.2, the net interest income is the excess of interest received over interest paid. It is the role of the asset-liability management function within the bank to ensure that the net interest margin, which is net interest income divided by income-producing assets, remains roughly constant through time. This section considers how this is done.

How can fluctuations in net interest margin occur? Consider a simple situation where a bank offers consumers a one-year and a five-year deposit rate as well as a one-year and five-year mortgage rate. The rates are shown in Table 14.6. We make the simplifying assumption that the expected one-year interest rate for future time periods equals the one-year rate prevailing in the market today. Loosely speaking, this means that market participants consider interest rate increases to be just as likely as interest rate decreases. As a result, the rates in Table 14.6 are “fair” in that they reflect the market’s expectations. Investing money for one year and reinvesting for four further one-year periods leads to an uncertain return. But, given our assumptions, the expected overall return is the same as a single five-year investment. Similarly, borrowing money for one year and refinancing each year for the next four years leads to the same expected financing costs as a single five-year loan.

Suppose you have money to deposit and agree with the prevailing view that interest rate increases are just as likely as interest rate decreases. Would you choose to deposit your money for one year at 3% per annum or for five years at 3% per annum? The chances are that you would choose one year because this gives you more financial flexibility. It ties up your funds for a shorter period of time.

Now suppose that you want a mortgage. Again, you agree with the prevailing view that interest rate increases are just as likely as interest rate decreases. Would you choose a one-year mortgage at 6% or a five-year mortgage at 6%? The chances are that you would choose a five-year mortgage because it fixes your borrowing rate for the next five years and subjects you to less refinancing risk.

When the bank posts the rates shown in Table 14.6, it is likely to find that the majority of its depositors opt for a one-year maturity and the majority of the customers seeking mortgages opt for a five-year maturity. This creates an asset/liability mismatch for the bank and subjects its net interest income to risks. The deposits that are financing

Table 14.6 Example of Rates Offered by a Bank to Its Customers

Maturity (years)	Deposit Rate	Mortgage Rate
1	3%	6%
5	3%	6%

Table 14.7 Five-Year Rates Are Increased in an Attempt to Match Maturities of Assets and Liabilities

Maturity (years)	Deposit Rate	Mortgage Rate
1	3%	6%
5	4%	7%

the five-year 6% mortgages are rolled over every year. There is no problem if interest rates fall. After one year, the bank will find itself financing the five-year 6% mortgages with deposits that cost less than 3% and net interest income will increase. However, if interest rates rise, the deposits that are financing the 6% mortgages will cost more than 3% and net interest income will decline. Suppose that there is a 3% rise in interest rates during the first two years. This would reduce net interest income for the third year to zero.

It is the job of the asset-liability management group to ensure that this type of interest rate risk is minimized. One way of doing this is to ensure that the maturities of the assets on which interest is earned and the maturities of the liabilities on which interest is paid are matched. In our example, the matching can be achieved by increasing the five-year rate on both deposits and mortgages. For example, the bank could move to the situation in Table 14.7 where the five-year deposit rate is 4% and the five-year mortgage rate is 7%. This would make five-year deposits relatively more attractive and one-year mortgages relatively more attractive. Some customers who chose one-year deposits when the rates were as in Table 14.6 will choose five-year deposits when rates are as in Table 14.7. Some customers who chose five-year mortgages when the rates were as in Table 14.6 will choose one-year mortgages. This may lead to the maturities of assets and liabilities being matched. If there is still an imbalance with depositors tending to choose a one-year maturity and borrowers a five-year maturity, five-year deposit and mortgage rates could be increased even further. Eventually the imbalance will disappear.

The net result of all banks behaving in the way we have just described is that long-term rates tend to be higher than those predicted by expected future short-term rates. This phenomenon is referred to as *liquidity preference theory*. It leads to long-term rates being higher than short-term rates most of the time. Even when the market expects a small decline in short-term rates, liquidity preference theory is likely to cause long-term rates to be higher than short-term rates. Only when a steep decline in interest rates is expected will long-term rates be less than short-term rates.

Many banks now have sophisticated systems for monitoring the decisions being made by customers so that, when they detect small differences between the maturities of the assets and liabilities being chosen, they can fine-tune the rates they offer. Often derivatives such as interest rate swaps are used to manage their exposures (see Example 5.1 in Chapter 5). The result of all this is that net interest margin is usually stable. This has not always been the case. In the 1980s in the United States, the failures of savings and loans companies were largely a result of their failure to match maturities for assets and liabilities.

14.4.1 Liquidity

In addition to eroding net interest margin, a mismatch of assets and liabilities can lead to liquidity problems. A bank that funds long-term loans with short-term deposits has to replace maturing deposits with new deposits on a regular basis. (This is sometimes referred to as *rolling over* the deposits.) If depositors lose confidence in the bank, it might find it difficult to do this. A well-known example of a financial institution that failed because of liquidity problems is Northern Rock in the United Kingdom. It chose to finance much of its mortgage portfolio with wholesale deposits, some lasting only three months. Starting in September 2007, the depositors became nervous because of the problems surfacing in the United States. As a result, Northern Rock was unable to finance its assets and was taken over by the UK government in early 2008 (see Business Snapshot 21.1 for more details). In the United States, Bear Stearns and Lehman Brothers experienced similar problems in rolling over their wholesale deposits.

Many of the problems during the Global Financial Crisis that started in 2007 were caused by a shortage of liquidity. As often happens during stressed market conditions, there was a flight to quality where investors looked for very safe investments and were not prepared to take credit risks. Bank regulators have now recognized the need to set liquidity requirements, as well as capital requirements, for banks. Chapter 21 discusses liquidity issues in more detail and Chapter 26 explains the liquidity requirements of regulators.

14.5 Duration

Duration is a widely used measure of a portfolio's exposure to yield curve movements. Suppose γ is a bond's yield and B is its market price. The duration D of the bond is defined as

$$D = -\frac{1}{B} \frac{\Delta B}{\Delta \gamma} \quad (14.4)$$

so that

$$\Delta B = -DB\Delta\gamma \quad (14.5)$$

where $\Delta\gamma$ is a small change in the bond's yield and ΔB is the corresponding change in its price. Duration measures the sensitivity of percentage changes in the bond's price to changes in its yield. Using calculus notation, we can write

$$D = -\frac{1}{B} \frac{dB}{d\gamma} \quad (14.6)$$

Consider a bond that provides cash flows $c_1, c_2, \dots, c_n$ at times $t_1, t_2, \dots, t_n$. (The cash flows consist of the coupon and principal payments on the bond.) The bond yield, γ , is defined as the discount rate that equates the bond's theoretical price to its market price. We denote the present value of the cash flow c_i , discounted from time t_i to today at rate γ , by v_i so that the price of the bond is

$$B = \sum_{i=1}^n v_i$$

An alternative definition of duration is

$$D = \sum_{i=1}^n t_i \left(\frac{v_i}{B} \right) \quad (14.7)$$

The term in parentheses in equation (14.7) is the ratio of the present value of the cash flow at time t_i to the bond price. Equation (14.7) therefore defines duration as a weighted average of the times when payments are made, with the weight applied to time t_i being equal to the proportion of the bond's total present value provided by the cash flow at time t_i . (The sum of the weights is 1.0.) This definition explains where the term duration comes from. Duration is a measure of how long the bondholder has to wait for cash flows. A zero-coupon bond that lasts n years has a duration of n years. However, a coupon-bearing bond lasting n years has a duration of less than n years, because the holder receives some of the cash payments prior to year n .

If the bond's yield, y , in equation (14.4) is measured with continuous compounding, it turns out that the definitions of duration in equations (14.4) and (14.7) are the same. (See Problem 14.15.)

Consider a three-year 10% coupon bond with a face value of \$100. Suppose that the yield on the bond is 12% per annum with continuous compounding. This means that $y = 0.12$. Coupon payments of \$5 are made every six months. Table 14.8 shows the calculations necessary to determine the bond's duration. The present values of the bond's cash flows, using the yield as the discount rate, are shown in column 3. (For example, the present value of the first cash flow is $5e^{-0.12 \times 0.5} = 4.709$.) The sum of the numbers in column 3 is the bond's market price, 94.213. The weights are calculated by dividing the numbers in column 3 by 94.213. The sum of the numbers in column 5 gives the duration as 2.653 years.

Small changes in interest rates are often measured in *basis points*. A basis point is 0.01% per annum. The following example shows that equation (14.5) is correct when duration is defined as in equation (14.7) and yields are measured with continuous compounding.

Table 14.8 Calculation of Duration

Time (years)	Cash Flow (\$)	Present Value	Weight	Time $\times$ Weight
0.5	5	4.709	0.050	0.025
1.0	5	4.435	0.047	0.047
1.5	5	4.176	0.044	0.066
2.0	5	3.933	0.042	0.083
2.5	5	3.704	0.039	0.098
3.0	105	73.256	0.778	2.333
Total	130	94.213	1.000	2.653

Example 14.2

For the bond in Table 14.8, the bond price, B , is 94.213 and the duration, D , is 2.653, so that equation (14.5) gives

$$\Delta B = -94.213 \times 2.653 \Delta y$$

or

$$\Delta B = -249.95 \Delta y$$

When the yield on the bond increases by 10 basis points ($= 0.1\%$), $\Delta y = +0.001$. The duration relationship predicts that $\Delta B = -249.95 \times 0.001 = -0.250$ so that the bond price goes down to $94.213 - 0.250 = 93.963$. How accurate is this? When the bond yield increases by 10 basis points to 12.1%, the bond price is

$$\begin{aligned} & 5e^{-0.121 \times 0.5} + 5e^{-0.121 \times 1.0} + 5e^{-0.121 \times 1.5} + 5e^{-0.121 \times 2.0} \\ & + 5e^{-0.121 \times 2.5} + 105e^{-0.121 \times 3.0} = 93.963 \end{aligned}$$

which is (to three decimal places) the same as that predicted by the duration relationship.

14.5.1 Modified Duration

The definition of duration in equation (14.7) was suggested by Frederick Macaulay in 1938. It is referred to as *Macaulay's duration*. As mentioned, when the yield y on the bond is measured with continuous compounding, it is equivalent to the definition in equation (14.4). When duration is defined using equation (14.4) and other compounding frequencies are used for y , a small adjustment is necessary to Macaulay's duration. When y is measured with annual compounding, it can be shown that the expression for D in equation (14.7) must be divided by $1 + y$. More generally, when y is expressed with a compounding frequency of m times per year, it must be divided by $1 + y/m$. (See Problem 14.15.) Duration defined with these adjustments to equation (14.7) is referred to as *modified duration*.

Example 14.3

The bond in Table 14.8 has a price of 94.213 and a duration of 2.653. The yield, expressed with semiannual compounding, is 12.3673%. (See Appendix A.) The (modified) duration appropriate for calculating sensitivity to the yield when it is expressed with semiannual compounding is

$$\frac{2.653}{1 + 0.123673/2} = 2.4985$$

From equation (14.5),

$$\Delta B = -94.213 \times 2.4985 \Delta y$$

or

$$\Delta B = -235.39 \Delta y$$

When the yield (semiannually compounded) increases by 10 basis points ($= 0.1\%$), $\Delta y = +0.001$. The duration relationship predicts that we expect ΔB to be $-235.39 \times 0.001 = -0.235$ so that the bond price goes down to $94.213 - 0.235 = 93.978$. How accurate is this? When the bond yield (semiannually compounded) increases by 10 basis points to 12.4673% (or to 12.0941% with continuous compounding), an exact calculation similar to that in the previous example shows that the bond price becomes 93.978. This shows that the modified duration is accurate for small yield changes.

14.5.2 Dollar Duration

The dollar duration of a bond is defined as the product of its duration and its price. If $D_{\$}$ is dollar duration, it follows from equation (14.5) that

$$\Delta B = -D_{\$}\Delta y$$

or using calculus notation

$$D_{\$} = -\frac{dB}{dy}$$

Whereas duration relates proportional changes in a bond's price to its yield, dollar duration relates actual changes in the bond's price to its yield. Dollar duration is analogous to the delta measure.

14.6 Convexity

The duration relationship measures exposure to small changes in yields. This is illustrated in Figure 14.3, which shows the relationship between the percentage change in value and change in yield for two bonds with the same duration. The gradients of the two curves are the same at the origin. This means that both portfolios change in value by the same percentage for small yield changes, as predicted by equation (14.5). For large yield changes, the bonds behave differently. Bond *X* has more curvature in its relationship with yields than bond *Y*. A factor known as *convexity* measures this curvature and can be used to improve the relationship between bond prices and yields.

The convexity for a bond is

$$C = \frac{1}{B} \frac{d^2 B}{dy^2} = \frac{\sum_{i=1}^n c_i t_i^2 e^{-\gamma t_i}}{B}$$

where γ is the bond's yield measured with continuous compounding. This is the weighted average of the square of the time to the receipt of cash flows. From Appendix G, a second order approximation to the change in the bond price is

$$\Delta B = \frac{dB}{dy} \Delta y + \frac{1}{2} \frac{d^2 B}{dy^2} \Delta y^2$$

This leads to

$$\frac{\Delta B}{B} = -D \Delta y + \frac{1}{2} C (\Delta y)^2 \quad (14.8)$$

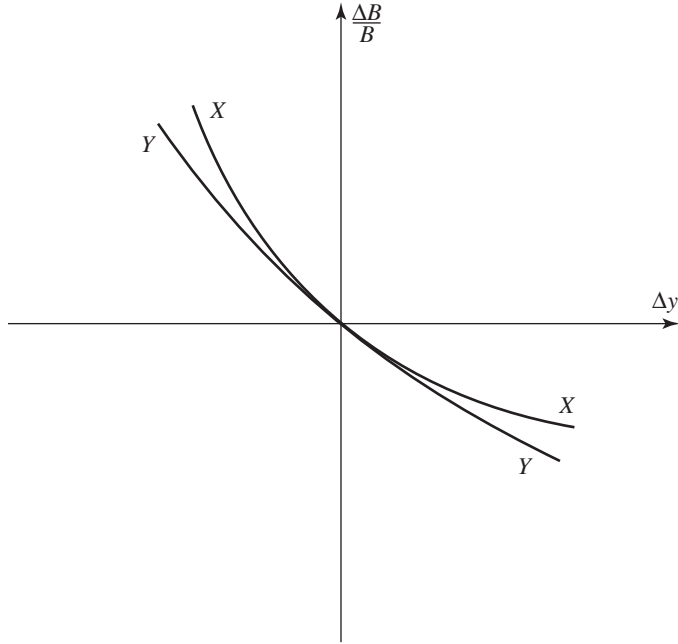


Figure 14.3 Two Bonds with the Same Duration and Different Convexities

Example 14.4

Consider again the bond in Table 14.8. The bond price, B , is 94.213, and the duration, D , is 2.653. The convexity is

$$0.05 \times 0.5^2 + 0.047 \times 1.0^2 + 0.044 \times 1.5^2 + 0.042 \times 2.0^2 + 0.039 \times 2.5^2 \\ + 0.779 \times 3.0^2 = 7.570$$

The convexity relationship in equation (14.8) is therefore

$$\frac{\Delta B}{B} = -2.653\Delta y + \frac{1}{2} \times 7.570 \times (\Delta y)^2$$

Consider a 2% change in the bond yield from 12% to 14%. The duration relationship predicts that the dollar change in the value of the bond will be $-94.213 \times 2.653 \times 0.02 = -4.999$. The convexity relationship predicts that it will be

$$-94.213 \times 2.653 \times 0.02 + 0.5 \times 94.213 \times 7.570 \times 0.02^2 = -4.856$$

The actual change in the value of the bond is -4.859 . This shows that the convexity relationship gives much more accurate results than duration for a large change in the bond yield.

14.6.1 Dollar Convexity

The *dollar convexity* of a bond, $C_{\$}$, can be defined analogously to dollar duration as the product of convexity and the value of the bond. This means that

$$C_{\$} = \frac{d^2 B}{dy^2}$$

and shows that dollar convexity is analogous to the gamma measure, which will be introduced in Chapter 15.

14.7 Generalization

So far we have used duration and convexity to measure the sensitivity of the price of a single bond to interest rates. The definitions of duration and convexity can be generalized so that they apply to a portfolio of bonds—or to any portfolio of interest-rate-dependent instruments. We define a parallel shift in the zero-coupon yield curve as a shift where all zero-coupon interest rates change by the same amount, as indicated in Figure 14.4.

Suppose that P is the value of the portfolio of interest-rate-dependent securities. We can make a small parallel shift in the zero-coupon yield curve and observe the change ΔP in P . Duration is defined as

$$D = -\frac{1}{P} \frac{\Delta P}{\Delta y} \quad (14.9)$$

where Δy is the size of the small parallel shift.³ Equation (14.9) is equivalent to

$$\frac{\Delta P}{P} = -D \Delta y \quad (14.10)$$

Suppose a portfolio consists of a number of interest-rate-dependent assets. The i th asset is worth X_i and has a duration D_i ($i = 1, 2, \dots, n$). Define ΔX_i as the change in

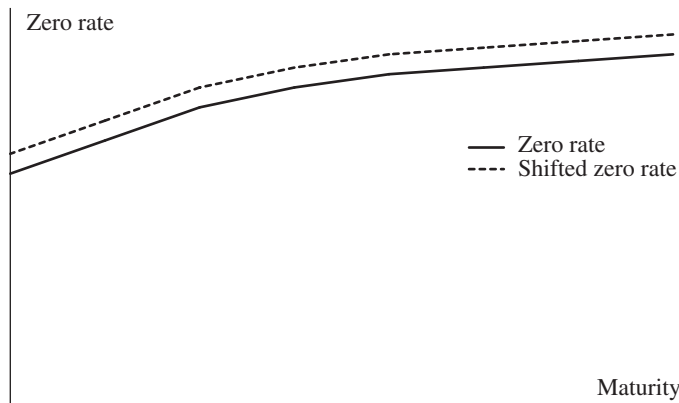


Figure 14.4 A Parallel Shift in Zero Rates

³A small parallel shift of Δy in the zero-coupon yield curve leads to the yield of all bonds changing by approximately Δy .

the value of X_i arising from the yield curve shift Δy . It follows that $P = \sum_{i=1}^n X_i$ and $\Delta P = \sum_{i=1}^n \Delta X_i$ so that from equation (14.9) the duration of the portfolio is given by

$$D = -\frac{1}{P} \sum_{i=1}^n \frac{\Delta X_i}{\Delta y}$$

The duration of the i th asset is

$$D_i = -\frac{1}{X_i} \frac{\Delta X_i}{\Delta y}$$

Hence

$$D = \sum_{i=1}^n \frac{X_i}{P} D_i$$

This shows that the duration D of a portfolio is the weighted average of the durations of the individual assets comprising the portfolio with the weight assigned to an asset being proportional to the value of the asset.

The dollar duration $D_{\$}$ of a portfolio can be defined as duration of the portfolio times the value of the portfolio:

$$D_{\$} = -\frac{\Delta P}{\Delta y}$$

This is a measure of the delta of the portfolio with respect to interest rates. The dollar duration of a portfolio consisting of a number of interest-rate-dependent assets is the sum of the dollar durations of the individual assets.

The convexity measure can be generalized in the same way as duration. For any interest-rate-dependent portfolio whose value is P we define the convexity C as $1/P$ times the second partial derivative of the value of the portfolio with respect to a parallel shift in the zero-coupon yield curve. Equation (14.8) is correct with B replaced by P :

$$\frac{\Delta P}{P} = -D\Delta y + \frac{1}{2}C(\Delta y)^2 \quad (14.11)$$

The relationship between the convexity of a portfolio and the convexity of the assets comprising the portfolio is similar to that for duration: the convexity of the portfolio is the weighted average of the convexities of the assets with the weights being proportional to the value of the assets. For a portfolio with a particular duration, the convexity tends to be greatest when the portfolio provides payments evenly over a long period of time. It is least when the payments are concentrated around one particular point in time.

The dollar convexity for a portfolio worth P can be defined as P times the convexity. This is a measure of the gamma of the portfolio with respect to interest rates. The dollar convexity of a portfolio consisting of a number of interest-rate-dependent positions is the sum of the dollar convexities of the individual assets.

14.7.1 Portfolio Immunization

A portfolio consisting of long and short positions in interest-rate-dependent assets can be protected against relatively small parallel shifts in the yield curve by ensuring that its

duration is zero. It can be protected against relatively large parallel shifts in the yield curve by ensuring that its duration and convexity are both zero or close to zero.

14.8 Nonparallel Yield Curve Shifts

Unfortunately the basic duration relationship in equation (14.10) only quantifies exposure to parallel yield curve shifts. The duration plus convexity relationship in equation (14.11) allows the shift to be relatively large, but it is still a parallel shift.

14.8.1 Partial Duration

Some researchers have extended duration measures so that nonparallel shifts can be considered. Reitano (1992) suggests a partial duration measure where just one point on the zero-coupon yield curve is shifted and all other points remain the same.⁴ Suppose that the zero curve is as shown in Table 14.9 and Figure 14.5. Shifting the five-year point involves changing the zero curve as indicated in Figure 14.6. In general, the partial duration of the portfolio for the i th point on the zero curve is

$$D_i = -\frac{1}{P} \frac{\Delta P_i}{\Delta y_i}$$

Table 14.9 Zero-Coupon Yield Curve (rates continuously compounded)

Maturity (years)	1	2	3	4	5	7	10
Rate (%)	4.0	4.5	4.8	5.0	5.1	5.2	5.3

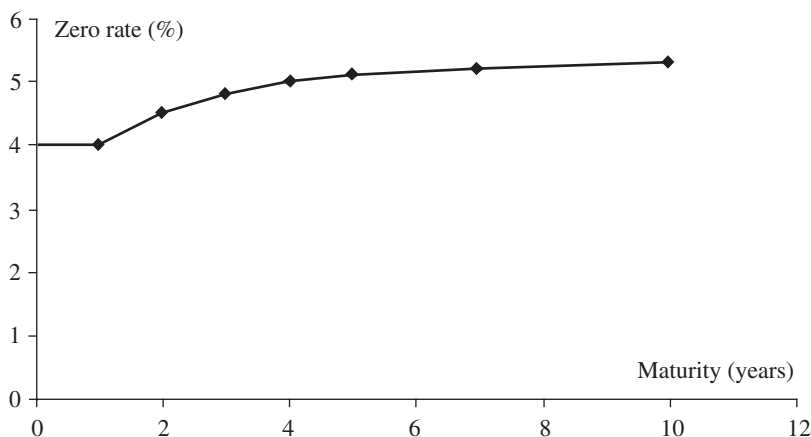


Figure 14.5 The Zero-Coupon Yield Curve (as shown in Table 14.9)

⁴See R. Reitano, "Nonparallel Yield Curve Shifts and Immunization," *Journal of Portfolio Management* (Spring 1992): 36–43.

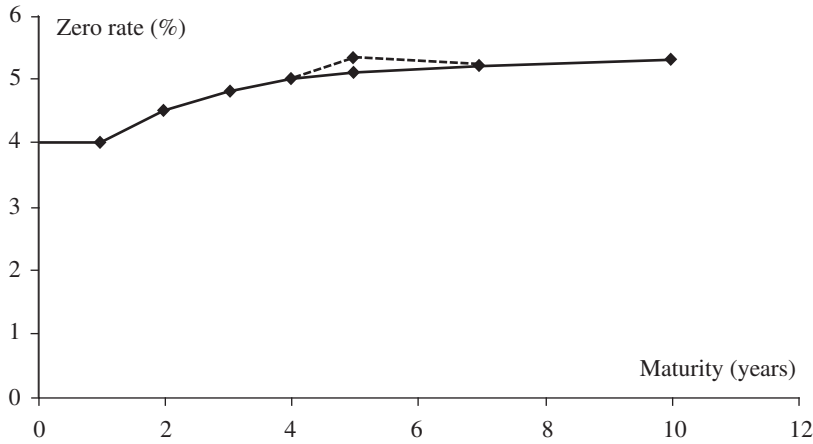


Figure 14.6 Change in Zero-Coupon Yield Curve When One Point Is Shifted

where Δy_i is a small change made to the i th point on the yield curve and ΔP_i is the resulting change in the portfolio value. The changes are calculated as explained in Section 14.2 so that the sum of the partial durations equals the usual duration measure.⁵ The percentage change in the portfolio value arising from Δy_i is $-D_i \Delta y_i$.

Suppose that the partial durations for a particular portfolio are as shown in Table 14.10. The duration of the portfolio (sum of the partial durations) is only 0.2. This means that the portfolio is relatively insensitive to parallel shifts in the yield curve. However, the durations for short maturities are positive while those for long maturities are negative. This means that the portfolio loses (gains) in value when short rates rise (fall). It gains (loses) in value when long rates rise (fall).

We are now in a position to go one step further and calculate the sensitivity of a portfolio value to any nonparallel shifts. Suppose that, in the case of the yield curve shown in Figure 14.5, we define a rotation where the changes to the 1-year, 2-year, 3-year, 4-year, 5-year, 7-year, and 10-year are $-3e$, $-2e$, $-e$, 0 , e , $3e$, and $6e$ for some small e . This is illustrated in Figure 14.7. From the partial durations in Table 14.10, the percentage change in the value of the portfolio arising from the rotation is

$$\begin{aligned}
 & -[0.2 \times (-3e) + 0.6 \times (-2e) + 0.9 \times (-e) + 1.6 \times 0 + 2.0 \times e \\
 & \quad -2.1 \times 3e - 3.0 \times 6e] = 25.0e
 \end{aligned}$$

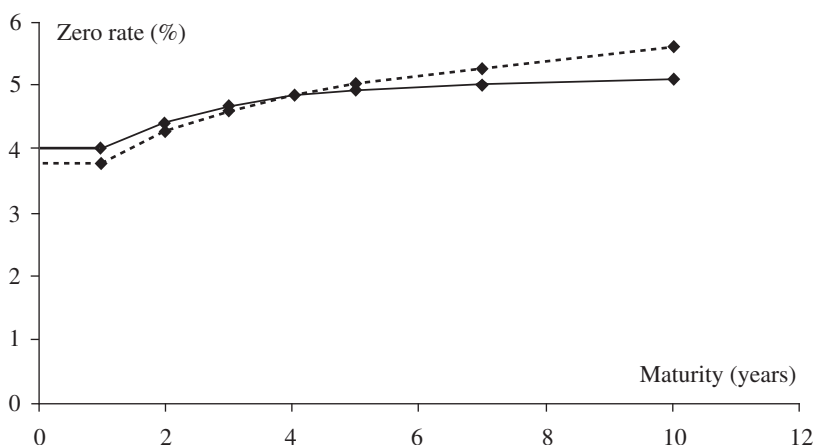
For a parallel shift of e in the yield curve, the percentage change in the value of the portfolio is $-0.2e$. This shows that a portfolio that gives rise to the partial durations in Table 14.10 is much more heavily exposed to a rotation of the yield curve than to a parallel shift.

In Section 14.8 we explained how the usual duration measure can be converted to a dollar duration measure by multiplying by the value of the portfolio. The same is true of

⁵When the i th point on the zero curve is shifted, the other points are not shifted and rates on the shifted yield curve are calculated using linear interpolation as indicated in Figure 14.6.

Table 14.10 Partial Durations for a Portfolio

Maturity (years)	1	2	3	4	5	7	10	Total
Duration	0.2	0.6	0.9	1.6	2.0	-2.1	-3.0	0.2

**Figure 14.7** A Rotation of the Yield Curve

partial duration measures. The result is deltas describing the rate of change in the value of the portfolio with respect to a small change in one vertex of the term structure.

14.8.2 Bucket Deltas

A variation on the partial duration approach is to divide the yield curve into a number of segments or buckets and calculate for each bucket the dollar impact of changing all the zero rates corresponding to the bucket by one basis point while keeping all other zero rates unchanged. This approach is often used in asset-liability management (see Section 14.4) and is referred to as *GAP management*. Figure 14.8 shows the type of change that would be considered for the segment of the zero curve between 2.0 and 3.0 years in Figure 14.5. The sum of the deltas for all the segments equals the dollar impact of a one-basis-point parallel shift (i.e., a DV01.)

14.8.3 Calculating Deltas to Facilitate Hedging

One of the problems with the exposure measures that we have considered so far is that they are not designed to make hedging easy. Consider the partial durations in Table 14.10. If we plan to hedge our portfolio with zero-coupon bonds, we can calculate the position in a one-year zero-coupon bond to zero out the exposure to the one-year rate, the position in a two-year zero-coupon bond to zero out the exposure to the two-year rate, and so on. But, if other instruments such as swaps or coupon-bearing bonds are used, a much more complicated analysis is necessary.

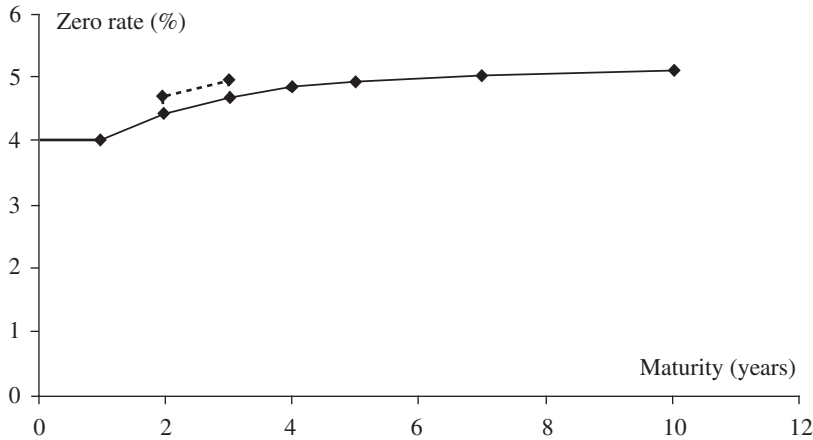


Figure 14.8 Change Considered to Yield Curve When Bucketing Approach Is Used

Often traders tend to use positions in the instruments that have been used to construct the zero curve to hedge their exposure. For example, a government bond trader is likely to take positions in the actively traded government bonds that were used to construct the Treasury zero curve when hedging. To facilitate hedging, traders therefore often calculate the impact of small changes in the quotes for each of the instruments used to construct the zero curve. The quote for the instrument is changed by a small amount, the zero-coupon yield curve is recomputed, and the portfolio revalued.

Summary

A financial institution has exposures to many different yield curves. To calculate VaR and ES, it can either regard each rate as a separate risk factor or use principal components analysis to find a small number of factors that describe movements in a yield curve. A bank's net interest margin is a measure of the excess of the interest rate it earns over the interest rate it pays. There are now well-established asset/liability management procedures to ensure that this remains roughly constant from year to year.

An important concept in interest rate markets is duration. Duration measures the sensitivity of the value of a portfolio to a small parallel shift in the zero-coupon yield curve. The relationship is

$$\Delta P = -PD\Delta y$$

where P is the value of the portfolio, D is the duration of the portfolio, Δy is the size of a small parallel shift in the zero curve, and ΔP is the resultant effect on the value of the portfolio. A more precise relationship is

$$\Delta P = -PD\Delta y + \frac{1}{2}PC(\Delta y)^2$$

where C is the convexity of the portfolio. This relationship is accurate for relatively large parallel shifts in the yield curve but does not quantify the exposure to nonparallel shifts.

To quantify exposure to all the different ways the yield curve can change through time, several duration or delta measures are necessary. There are a number of ways these can be defined. A principal components analysis can be a useful alternative to calculating multiple deltas. It shows that the yield curve shifts that occur in practice are, to a large extent, a linear sum of two or three standard shifts. If a portfolio manager is hedged against these standard shifts, he or she is therefore also well hedged against the shifts that occur in practice.

Further Reading

- Duffie, D. "Debt Management and Interest Rate Risk." In *Risk Management: Challenges and Solutions*, edited by W. Beaver and G. Parker. New York: McGraw-Hill, 1994.
- Fabozzi, F. J. *Bond Markets, Analysis and Strategies*. 8th ed. Upper Saddle River, NJ: Pearson, 2012.
- Jorion, P. *Big Bets Gone Bad: Derivatives and Bankruptcy in Orange County*. New York: Academic Press, 1995.
- Reitano, R. "Nonparallel Yield Curve Shifts and Immunization." *Journal of Portfolio Management* (Spring 1992): 36–43.

Practice Questions and Problems (Answers at End of Book)

- 14.1 Suppose that a bank has \$5 billion of one-year loans and \$20 billion of five-year loans. These are financed by \$15 billion of one-year deposits and \$10 billion of five-year deposits. Explain the impact on the bank's net interest income of interest rates increasing by 1% every year for the next three years.
- 14.2 Explain why long-term rates are higher than short-term rates most of the time. Under what circumstances would you expect long-term rates to be lower than short-term rates?
- 14.3 Describe two ways of handling interest-rate-dependent instruments when the model-building approach is used to calculate VaR.
- 14.4 Suppose that the value of a portfolio increases by \$50,000 for each one-basis-point increase in the 12-year rate and has no other sensitivities. The multiple-vertex approach is used to model with the following vertices: 3 months, 6 months, 1 year, 2 years, 3 years, 5 years, 10 years, 15 years, 20 years, and 30 years. What is the sensitivity of the portfolio to a one-basis-point increase in each vertex of the term structure?
- 14.5 A portfolio has exposure to the two-year interest rate and the five-year interest rate. A one-basis-point increase in the two-year rate causes the value of the portfolio to increase in value by \$10,000. A one-basis-point increase in the five-year rate causes the value of the portfolio to increase by \$8,000. The standard deviation per day of the two-year rate and that of the five-year rate are 7 and 8 basis points,

- respectively, and the correlation between the two rates is 0.8. What is the portfolio's expected shortfall when the confidence level is 98% and the time horizon is five days?
- 14.6 What does duration tell you about the sensitivity of a bond portfolio to interest rates? What are the limitations of the duration measure?
- 14.7 A five-year bond with a yield of 11% (continuously compounded) pays an 8% coupon at the end of each year.
- (a) What is the bond's price?
 - (b) What is the bond's duration?
 - (c) Use the duration to calculate the effect on the bond's price of a 0.2% decrease in its yield.
 - (d) Recalculate the bond's price on the basis of a 10.8% per annum yield and verify that the result is in agreement with your answer to (c).
- 14.8 Repeat Problem 14.7 on the assumption that the yield is compounded annually. Use modified durations.
- 14.9 A six-year bond with a continuously compounded yield of 4% provides a 5% coupon at the end of each year. Use duration and convexity to estimate the effect of a 1% increase in the yield on the price of the bond. How accurate is the estimate?
- 14.10 Explain three ways in which multiple deltas can be calculated to manage nonparallel yield curve shifts.
- 14.11 Why are partial durations potentially more useful to an analyst than the usual duration measure?
- 14.12 Use the partial durations in Table 14.10 to calculate the impact of a shift in the yield curve on a \$10 million portfolio where the 1-, 2-, 3-, 4-, 5-, 7-, and 10-year rates increase by 10, 8, 7, 6, 5, 3, and 1 basis points, respectively.
- 14.13 How are "dollar duration" and "dollar convexity" defined?
- 14.14 What is the relationship between (a) the duration, (b) the partial durations, and (c) the DV01 of a portfolio?

Further Questions

- 14.15 Prove (a) that the definitions of duration in equations (14.4) and (14.7) are the same when y is continuously compounded and (b) that when y is compounded m times per year they are the same if the right hand side of equation (14.7) is divided by $1 + y/m$.
- 14.16 Suppose that a bank has \$10 billion of one-year loans and \$30 billion of five-year loans. These are financed by \$35 billion of one-year deposits and \$5 billion of five-year deposits. The bank has equity totaling \$2 billion and its return on equity is currently 12%. Estimate what change in interest rates next year would lead to the bank's return on equity being reduced to zero. Assume that the bank is subject to a tax rate of 30%.

- 14.17 Portfolio A consists of a one-year zero-coupon bond with a face value of \$2,000 and a 10-year zero-coupon bond with a face value of \$6,000. Portfolio B consists of a 5.95-year zero-coupon bond with a face value of \$5,000. The current yield on all bonds is 10% per annum (continuously compounded).
- (a) Show that both portfolios have the same duration.
 - (b) Show that the percentage changes in the values of the two portfolios for a 0.1% per annum increase in yields are the same.
 - (c) What are the percentage changes in the values of the two portfolios for a 5% per annum increase in yields?
- 14.18 What are the convexities of the portfolios in Problem 14.17? To what extent do (a) duration and (b) convexity explain the difference between the percentage changes calculated in part (c) of Problem 14.17?
- 14.19 When the partial durations are as in Table 14.10, estimate the effect of a shift in the yield curve where the ten-year rate stays the same, the one-year rate moves up by 9e, and the movements in intermediate rates are calculated by interpolation between 9e and 0. How could your answer be calculated from the results for the rotation calculated in Section 14.8?
- 14.20 Suppose that the change in a portfolio value for a one-basis-point shift in the 1-, 2-, 3-, 5-, 7-, 10-, 20-, and 30-year rates are (in \$ million) +5, -3, -1, +2, +5, +7, +8, and +1, respectively. Estimate the delta of the portfolio with respect to the first three factors in Table 14.3. Quantify the relative importance of the three factors for this portfolio.
- 14.21 The text calculates a VaR estimate for the example in Table 14.5 assuming two factors. How does the estimate change if you assume (a) one factor and (b) three factors?
- 14.22 A company has a long position in a two-year bond and a three-year bond, as well as a short position in a five-year bond. Each bond has a principal of \$100 million and pays a 5% coupon annually. Calculate the company's exposure to the one-year, two-year, three-year, four-year, and five-year rates.
- 14.23 A company has a position in bonds worth \$6 million. The modified duration of the portfolio is 5.2 years. Assume that only parallel shifts in the yield curve can take place and that the standard deviation of the daily yield change (when yield is measured in percent) is 0.09. Use the duration model to estimate the 20-day 90% VaR for the portfolio. Explain carefully the weaknesses of this approach to calculating VaR. Explain two alternatives that give more accuracy.

Chapter 15

Derivatives Risk

So far most of our discussion has involved products where the price is linearly dependent on market variables. We now move on to consider derivatives, which are products where the linearity assumption does not hold. We explain what are termed the “Greek letters,” often referred to simply as the “Greeks,” which are used to manage the risk in derivatives positions. Each of the Greeks measures a different aspect of risk. Traders calculate their Greeks at the end of each day and are required to take action if the internal risk limits of the financial institution they work for are exceeded. Failure to take this action is liable to lead to immediate dismissal. Using historical simulation to calculate value at risk or expected shortfall for derivatives portfolios is straightforward. Using the model-building approach is more complicated. We will discuss some approximations that can be used.

15.1 Delta

Imagine that you are a trader working for a U.S. bank and are responsible for all trades involving gold. The current price of gold is \$1,300 per ounce. Table 15.1 shows a summary of your portfolio (known as your “book”). How can you manage your risks?

The value of your portfolio is currently $-\$5,683,000$. (This could be partly because you have been a net seller of options and partly because the market has moved against you.) One way of investigating the risks you face is to revalue the portfolio on the assumption that there is a small increase in the price of gold from \$1,300 per ounce to \$1,300.10 per ounce. Suppose that this \$0.10 increase in the price of gold decreases

Table 15.1 Summary of Gold Portfolio

Position	Value (\$)
Spot gold	3,180,000
Forward contracts	−3,060,000
Futures contracts	2,000
Swaps	180,000
Options	−6,110,000
Exotics	125,000
Total	−5,683,000

the value of your portfolio by \$100 from −\$5,683,000 to −\$5,683,100. This means that the sensitivity of the portfolio to the price of gold is

$$\frac{-100}{0.1} = -1,000$$

This is referred to as the *delta* of the portfolio (It is a Greek letter we have used in earlier chapters.). The portfolio loses value at a rate of about \$1,000 per \$1 increase in the price of gold. Similarly, it gains value at a rate of about \$1,000 per \$1 decrease in the price of gold.

In general, the *delta* of a portfolio with respect to a market variable is

$$\frac{\Delta P}{\Delta S}$$

where ΔS is a small increase in the value of the variable and ΔP is the resulting change in the value of the portfolio. Using calculus terminology, delta is the partial derivative of the portfolio value with respect to the value of the variable:

$$\text{Delta} = \frac{\partial P}{\partial S}$$

In our example, the trader can eliminate the delta exposure by buying 1,000 ounces of gold. This is because the delta of a long position in 1,000 ounces of gold is 1,000. (The position gains value at the rate of \$1,000 per \$1 increase in the price of gold.) This is known as *delta hedging*. When the hedging trade is combined with the existing portfolio the resultant portfolio has a delta of zero. Such a portfolio is referred to as *delta neutral*.

15.1.1 Linear Products

A linear product is one whose value at any given time is linearly dependent on the value of an underlying market variable (see Figure 15.1). If the underlying variable is the price of an asset such as gold, a spot position in the asset is clearly a linear product. The value of the position varies linearly with the price of the asset. A forward contract is also a linear product. However, options and other more complex derivative products are not linear products.

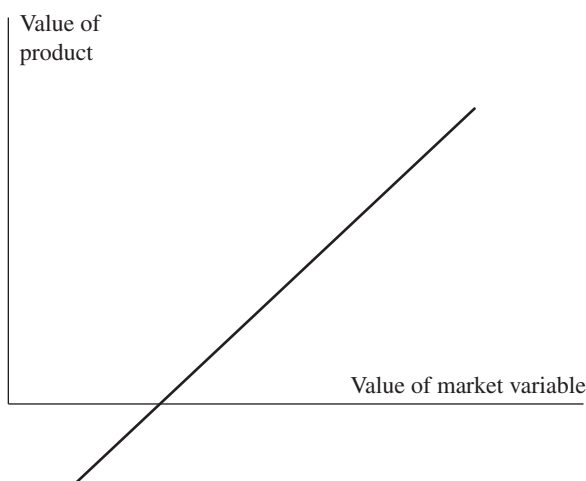


Figure 15.1 A Linear Product

A linear product can be hedged relatively easily. As a simple example, consider a U.S. bank that has entered into a forward contract with a corporate client where it agreed to sell the client 1 million euros for \$1.3 million in one year. Assume that the euro and dollar interest rates are 4% and 3% with annual compounding. This means that the present value of a 1 million euro cash flow in one year is $1,000,000/1.04 = 961,538$ euros. The present value of 1.3 million dollars in one year is $1,300,000/1.03 = 1,262,136$ dollars. Suppose that S is the value of one euro in dollars today. The value of the contract today in dollars is¹

$$1,262,136 - 961,538S$$

This shows that the value of the contract is linearly related to the exchange rate, S . The delta of the contract is $-961,538$. It can be hedged by buying 961,538 euros. Because of the linearity, the hedge provides protection against both small and large movements in S .

When the bank enters into the opposite transaction and agrees to buy one million euros in one year, the value of the contract is also linear in S

$$961,538S - 1,262,136$$

The bank has a delta of $+961,538$. It must hedge by shorting 961,538 euros. It does this by borrowing the euros today at 4% and immediately converting them to U.S. dollars. The one million euros received in one year are used to repay the loan.

¹See Appendix C for more information on the valuation of forward contracts.

BUSINESS SNAPSHOT 15.1

Hedging by Gold Mining Companies

It is natural for a gold mining company to consider hedging against changes in the price of gold. Typically it takes several years to extract all the gold from a mine. Once a gold mining company decides to go ahead with production at a particular mine, it has a big exposure to the price of gold. Indeed a mine that looks profitable at the outset could become unprofitable if the price of gold plunges.

Gold mining companies are careful to explain their hedging strategies to potential shareholders. Some gold mining companies do not hedge. They tend to attract shareholders who buy gold stocks because they want to benefit when the price of gold increases and are prepared to accept the risk of a loss from a decrease in the price of gold. Other companies choose to hedge. They estimate the number of ounces they will produce each month for the next few years and enter into futures or forward contracts to lock in the price that will be received.

Suppose you are Goldman Sachs and have just entered into a forward contract with a gold mining company where you agree to buy at a future time a large amount of gold at a fixed price. How do you hedge your risk? The answer is that you borrow gold from a central bank and sell it at the current market price. (The central banks of some of the countries that hold large amounts of gold are prepared to lend gold for a fee known as the gold lease rate.) At the end of the life of the forward contract, you buy gold from the gold mining company under the terms of the forward contract and use it to repay the central bank.

Shorting assets to hedge forward contracts is sometimes tricky. Gold is an interesting case in point. Financial institutions often find that they enter into very large forward contracts to buy gold from gold producers. This means that they need to borrow large quantities of gold to create a short position for hedging. As outlined in Business Snapshot 15.1, central banks are the source of the borrowed gold. A fee known as the gold lease rate is charged by central banks for lending the gold.

Linear products have the attractive property that hedges protect against large changes as well as small ones in the value of the underlying asset. They also have another related attractive property: the hedge, once it has been set up, never needs to be changed. (This is sometimes referred to as the “hedge and forget” property.) For an illustration of this, consider again the first forward contract we considered where a bank agrees to sell a client 1 million euros for 1.3 million dollars. A total of 961,538 euros are purchased to hedge the position. These can be invested at 4% for one year so that they grow to exactly 1 million euros in one year. This is exactly what the bank needs to complete the forward transaction in one year so that there is no need to adjust the hedge during the year.

15.1.2 Nonlinear Products

Options and other more complex derivatives dependent on the price of an underlying asset are nonlinear products. The relationship between the value of the product and the underlying asset price at any given time is nonlinear. This nonlinearity makes them more difficult to hedge for two reasons. First, making a nonlinear portfolio delta neutral only protects against small movements in the price of the underlying asset. Second, we are not in a hedge-and-forget situation. The hedge needs to be changed frequently. This is known as *dynamic hedging*.

Consider as an example a trader who sells 100,000 European call options on a non-dividend-paying stock when

1. Stock price is \$49
2. Strike price is \$50
3. Risk-free interest rate is 5%
4. Stock price volatility is 20% per annum
5. Time to option maturity is 20 weeks

We suppose that the amount received for the options is \$300,000 and that the trader has no other positions dependent on the stock.

The value of one option as a function of the underlying stock price is shown in Figure 15.2. (The risk-neutral valuation principle discussed in Chapter 10 shows that we do not need the expected return on the stock to value the option.) Because the function is nonlinear (i.e., curved), the delta of a position in one option depends on the stock price. When the stock price is low, the value of the option is almost zero and the dollar increase in the value of the option as a result of a small change in the stock price is also close to zero. This means that the delta of the option is close to zero. When the stock

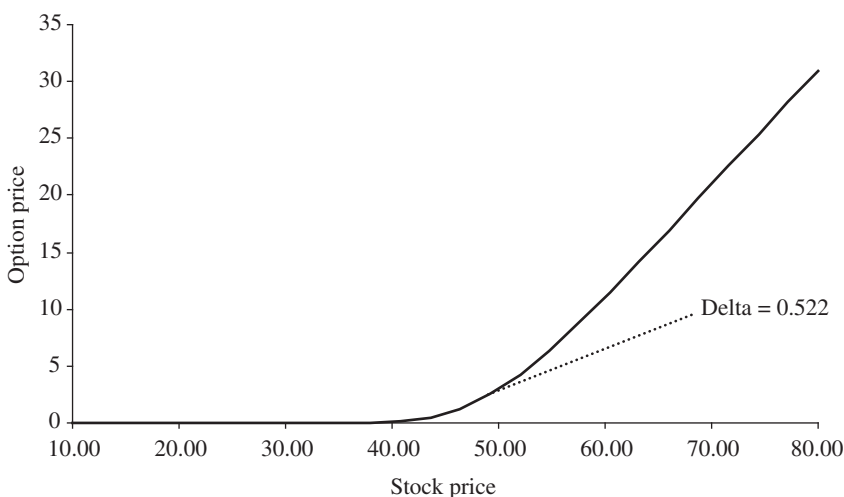


Figure 15.2 Value of Call Option as a Function of Stock Price (strike price = \$50, risk-free rate = 5%, volatility = 20%, time to maturity = 20 weeks)

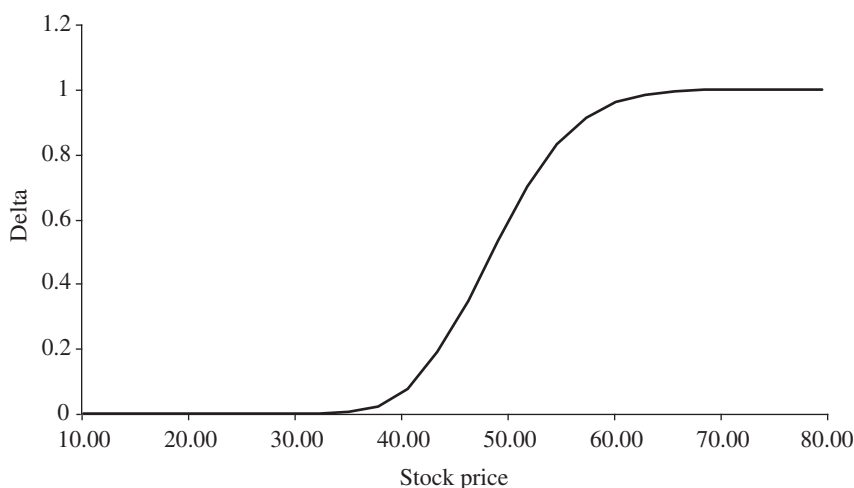


Figure 15.3 Delta of Call Option as a Function of Stock Price (strike price = \$50, risk-free rate = 5%, volatility = 20%, time to maturity = 20 weeks)

price is high, say \$70, the option is almost certain to be exercised and the delta of a position in one option is close to 1.0. This is because an increase (decrease) of ΔS in the stock price increases (decreases) the value of the option by almost ΔS .

In our example, the stock price is \$49. The delta of the option is the gradient of the curve at this point. This is 0.522, as indicated in Figure 15.2. (See Appendix E for the calculation of the Greeks for European options.) The delta of one option changes with the stock price in the way shown in Figure 15.3.² At the time of the trade, the value of an option to buy one share of the stock is \$2.40 and the delta of the option is 0.522. Because the trader is short 100,000 options, the value of the trader's portfolio is $-\$240,000$ and the delta of the portfolio is $-52,200$. The trader can feel pleased that the options have been sold for \$60,000 more than their theoretical value, but is faced with the problem of hedging the risk in the portfolio.

Immediately after the trade, the trader's portfolio can be made delta neutral by buying 52,200 shares of the underlying stock. If there is a small decrease (increase) in the stock price, the gain (loss) to the trader of the short option position should be offset by the loss (gain) on the shares. For example, if the stock price increases from \$49 to \$49.10, the value of the option position will decrease by about $52,200 \times 0.10 = \$5,220$, while the value of the shares will increase by this amount.

As mentioned, in the case of linear products, once the hedge has been set up it does not need to be changed. This is not the case for nonlinear products. To preserve delta neutrality, the hedge has to be adjusted periodically. This is known as *rebalancing*.

Tables 15.2 and 15.3 provide two examples of how rebalancing might work in our example. Rebalancing is assumed to be done weekly. As mentioned, the initial value of

²Figures 15.2 and 15.3 were produced with the RMFI software that can be downloaded from the author's website. The Black–Scholes–Merton model is selected by choosing “Black–Scholes–European” as the option type.

Table 15.2 Simulation of Delta Hedging (option closes in-the-money and cost of hedging is \$263,300)

Week	Stock Price	Delta	Shares Purchased	Cost of Shares Purchased (\$000)	Cumulative Cash Outflow (\$000)	Interest Cost (\$000)
0	49.00	0.522	52,200	2,557.8	2,557.8	2.5
1	48.12	0.458	(6,400)	(308.0)	2,252.3	2.2
2	47.37	0.400	(5,800)	(274.7)	1,979.8	1.9
3	50.25	0.596	19,600	984.9	2,966.6	2.9
4	51.75	0.693	9,700	502.0	3,471.5	3.3
5	53.12	0.774	8,100	430.3	3,905.1	3.8
6	53.00	0.771	(300)	(15.9)	3,893.0	3.7
7	51.87	0.706	(6,500)	(337.2)	3,559.5	3.4
8	51.38	0.674	(3,200)	(164.4)	3,398.5	3.3
9	53.00	0.787	11,300	598.9	4,000.7	3.8
10	49.88	0.550	(23,700)	(1,182.2)	2,822.3	2.7
11	48.50	0.413	(13,700)	(664.4)	2,160.6	2.1
12	49.88	0.542	12,900	643.5	2,806.2	2.7
13	50.37	0.591	4,900	246.8	3,055.7	2.9
14	52.13	0.768	17,700	922.7	3,981.3	3.8
15	51.88	0.759	(900)	(46.7)	3,938.4	3.8
16	52.87	0.865	10,600	560.4	4,502.6	4.3
17	54.87	0.978	11,300	620.0	5,126.9	4.9
18	54.62	0.990	1,200	65.5	5,197.3	5.0
19	55.87	1.000	1,000	55.9	5,258.2	5.1
20	57.25	1.000	0	0.0	5,263.3	

delta for a single option is 0.522 and the delta of the portfolio is $-52,200$. This means that, as soon as the option is written, \$2,557,800 must be borrowed to buy 52,200 shares at a price of \$49. The rate of interest is 5%. An interest cost of approximately \$2,500 is therefore incurred in the first week.

In Table 15.2, the stock price falls by the end of the first week to \$48.12. The delta declines to 0.458. A long position in 45,800 shares is now required to hedge the option position. A total of 6,400 ($= 52,200 - 45,800$) shares are therefore sold to maintain the delta neutrality of the hedge. The strategy realizes \$308,000 in cash, and the cumulative borrowings at the end of week 1 are reduced to \$2,252,300. During the second week the stock price reduces to \$47.37 and delta declines again. This leads to 5,800 shares being sold at the end of the second week. During the third week, the stock price increases to over \$50 and delta increases. This leads to 19,600 shares being purchased at the end of the third week. Toward the end of the life of the option, it becomes apparent that the option will be exercised and delta approaches 1.0. By week 20, therefore, the hedger owns 100,000 shares. The hedger receives \$5 million ($= 100,000 \times \50) for these shares when the option is exercised so that the total cost of writing the option and hedging it is \$263,300.

Table 15.3 Simulation of Delta Hedging (option closes out-of-the-money and cost of hedging = \$256,600)

Week	Stock Price	Delta	Shares Purchased	Cost of Shares Purchased (\$000)	Cumulative Cash Outflow (\$000)	Interest Cost (\$000)
0	49.00	0.522	52,200	2,557.8	2,557.8	2.5
1	49.75	0.568	4,600	228.9	2,789.2	2.7
2	52.00	0.705	13,700	712.4	3,504.3	3.4
3	50.00	0.579	(12,600)	(630.0)	2,877.7	2.8
4	48.38	0.459	(12,000)	(580.6)	2,299.9	2.2
5	48.25	0.443	(1,600)	(77.2)	2,224.9	2.1
6	48.75	0.475	3,200	156.0	2,383.0	2.3
7	49.63	0.540	6,500	322.6	2,707.9	2.6
8	48.25	0.420	(12,000)	(579.0)	2,131.5	2.1
9	48.25	0.410	(1,000)	(48.2)	2,085.4	2.0
10	51.12	0.658	24,800	1,267.8	3,355.2	3.2
11	51.50	0.692	3,400	175.1	3,533.5	3.4
12	49.88	0.542	(15,000)	(748.2)	2,788.7	2.7
13	49.88	0.538	(400)	(20.0)	2,771.4	2.7
14	48.75	0.400	(13,800)	(672.7)	2,101.4	2.0
15	47.50	0.236	(16,400)	(779.0)	1,324.4	1.3
16	48.00	0.261	2,500	120.0	1,445.7	1.4
17	46.25	0.062	(19,900)	(920.4)	526.7	0.5
18	48.13	0.183	12,100	582.4	1,109.6	1.1
19	46.63	0.007	(17,600)	(820.7)	290.0	0.3
20	48.12	0.000	(700)	(33.7)	256.6	

Table 15.3 illustrates an alternative sequence of events where the option closes out-of-the-money. As it becomes clear that the option will not be exercised, delta approaches zero. By week 20, the hedger therefore has no position in the underlying stock. The total costs incurred are \$256,600.

In Tables 15.2 and 15.3, the costs of hedging the option, when discounted to the beginning of the period, are close to, but not exactly, the same as the theoretical (Black–Scholes–Merton) price of \$240,000. If the hedging scheme worked perfectly, the cost of hedging would, after discounting, be exactly equal to the Black–Scholes–Merton price for every simulated stock price path. The reason for the variation in the cost of delta hedging is that the hedge is rebalanced only once a week. As rebalancing takes place more frequently, the variation in the cost of hedging is reduced. Of course, the examples in Tables 15.2 and 15.3 are idealized in that they assume the model underlying the Black–Scholes–Merton formula is exactly correct and there are no transactions costs.

Delta hedging aims to keep the value of the financial institution's position as close to unchanged as possible. Initially, the value of the written option is \$240,000. In the situation depicted in Table 15.2, the value of the option can be calculated as \$414,500

in week 9. Thus, the financial institution has lost $414,500 - 240,000 = \$174,500$ on its short option position. Its cash position, as measured by the cumulative cost, is \$1,442,900 worse in week 9 than in week 0. The value of the shares held has increased from \$2,557,800 to \$4,171,100 for a gain of \$1,613,300. The net effect is that the value of the financial institution's position has changed by only \$4,100 during the nine-week period.

15.1.3 *Where the Cost Comes From*

The delta-hedging procedure in Tables 15.2 and 15.3 in effect creates a long position in the option synthetically to neutralize the trader's short option position. As the tables illustrate, the procedure tends to involve selling stock just after the price has gone down and buying stock just after the price has gone up. It might be termed a buy-high, sell-low trading strategy! The cost of \$240,000 comes from the average difference between the price paid for the stock and the price realized for it.

15.1.4 *Economies of Scale*

Maintaining a delta-neutral position in a single option and the underlying asset, in the way that has just been described, is liable to be prohibitively expensive because of the transactions costs incurred on trades. Maintaining delta neutrality is more feasible for a large portfolio of derivatives dependent on a single asset because only one trade in the underlying asset is necessary to zero out delta for the whole portfolio. The hedging transactions costs are absorbed by the profits on many different trades. This shows that there are economies of scale in trading derivatives. It is not surprising that the derivatives market is dominated by a small number of large dealers.

15.2 Gamma

As mentioned, for a nonlinear portfolio, delta neutrality only provides protection against small changes in the price of the underlying asset.

The *gamma*, Γ , of a portfolio measures the extent to which large changes cause problems. Gamma is the rate of change of the portfolio's delta with respect to the price of the underlying asset. It is the second partial derivative of the portfolio with respect to asset price:

$$\text{Gamma} = \frac{\partial^2 P}{\partial S^2}$$

If gamma is small, delta changes slowly, and adjustments to keep a portfolio delta neutral need to be made only relatively infrequently. However, if gamma is large in absolute terms, delta is highly sensitive to the price of the underlying asset. It is then quite risky to leave a delta-neutral portfolio unchanged for any length of time. Figure 15.4 illustrates this point for an option on a stock. When the stock price moves from S to S' , delta hedging assumes that the option price moves from C to C' , when in fact it moves

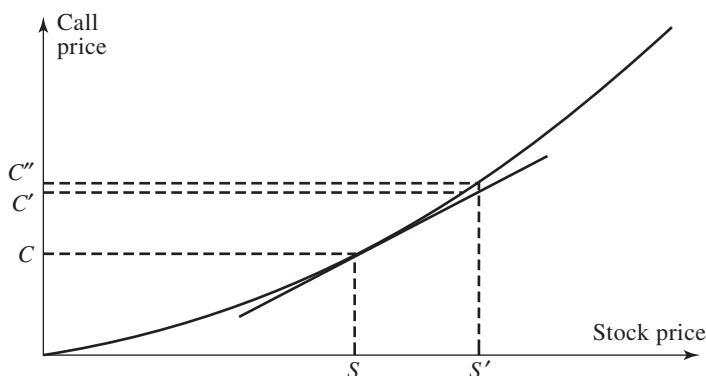


Figure 15.4 Hedging Error Introduced by Nonlinearity

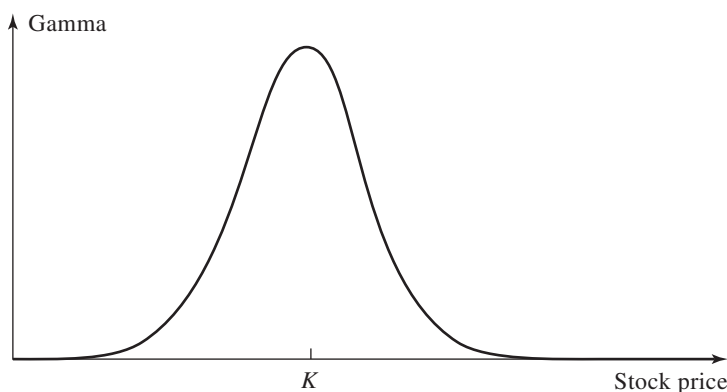


Figure 15.5 Relationship between Gamma of an Option and Price of Underlying Stock Where K Is the Option's Strike Price

from C to C'' . The difference between C' and C'' leads to a hedging error. This error depends on the curvature of the relationship between the option price and the stock price. Gamma measures this curvature.³

Gamma is positive for a long position in an option. The general way in which gamma varies with the price of the underlying stock is shown in Figure 15.5. Gamma is greatest for options where the stock price is close to the strike price K .

15.2.1 Making a Portfolio Gamma Neutral

A linear product has zero gamma and cannot be used to change the gamma of a portfolio. What is required is a position in an instrument, such as an option, that is not linearly dependent on the underlying asset price.

Suppose that a delta-neutral portfolio has a gamma equal to Γ , and a traded option has a gamma equal to Γ_T . If the number of traded options added to the portfolio is w_T , the gamma of the portfolio is

$$w_T \Gamma_T + \Gamma$$

³Indeed, the gamma of an option is sometimes referred to as its *curvature*.

Hence, the position in the traded option necessary to make the portfolio gamma neutral is $w_T = -\Gamma/\Gamma_T$. Including the traded option is likely to change the delta of the portfolio, so the position in the underlying asset then has to be changed to maintain delta neutrality. Note that the portfolio is gamma neutral only for a short period of time. As time passes, gamma neutrality can be maintained only if the position in the traded option is adjusted so that it is always equal to $-\Gamma/\Gamma_T$.

Making a delta-neutral portfolio gamma neutral can be regarded as a first correction for the fact that the position in the underlying asset cannot be changed continuously when delta hedging is used. Delta neutrality provides protection against relatively small asset price moves between rebalancing. Gamma neutrality provides protection against larger movements in the asset price between hedge rebalancing. Suppose that a portfolio is delta neutral and has a gamma of $-3,000$. The delta and gamma of a particular traded call option are 0.62 and 1.50 , respectively. The portfolio can be made gamma neutral by including in the portfolio a long position of

$$\frac{3,000}{1.5} = 2,000$$

in the call option. (The gamma of the portfolio is then $-3,000 + 1.5 \times 2,000 = 0$.) However, the delta of the portfolio will then change from zero to $2,000 \times 0.62 = 1,240$. A quantity, $1,240$, of the underlying asset must therefore be sold to keep it delta neutral.

15.3 Vega

Another source of risk in derivatives trading is the possibility that the volatility of the price of the underlying asset will change. In option valuation models, volatilities are often assumed to be constant, but in practice they do change through time. Spot positions and forwards do not depend on the volatility of underlying asset prices, but the prices of options and more complex derivatives change because of movements in volatility as well as because of changes in the asset price and the passage of time.

The *vega* of a portfolio, V , is the rate of change of the value of the portfolio with respect to the volatility, σ , of the underlying asset price.⁴

$$V = \frac{\partial P}{\partial \sigma}$$

If vega is high in absolute terms, the portfolio's value is very sensitive to small changes in volatility. If vega is low in absolute terms, volatility changes have relatively little impact on the value of the portfolio.

The vega of a portfolio can be changed by adding a position in a traded option. If V is the vega of the portfolio and V_T is the vega of a traded option, a position of $-V/V_T$

⁴Vega is the name given to one of the "Greek letters" in option pricing, but it is not one of the letters in the Greek alphabet.

in the traded option makes the portfolio instantaneously vega neutral. Unfortunately, a portfolio that is gamma neutral will not, in general, be vega neutral, and vice versa. If a hedger requires a portfolio to be both gamma and vega neutral, at least two traded derivatives dependent on the underlying asset must usually be used.

Example 15.1

Consider a portfolio dependent on the price of a single asset that is delta neutral, with a gamma of $-5,000$ and a vega of $-8,000$. The options shown in the table below can be traded. The portfolio could be made vega neutral by including a long position in 4,000 of Option 1. This would increase delta to 2,400 and require that 2,400 units of the asset be sold to maintain delta neutrality. The gamma of the portfolio would change from $-5,000$ to $-3,000$.

	Delta	Gamma	Vega
Portfolio	0	$-5,000$	$-8,000$
Option 1	0.6	0.5	2.0
Option 2	0.5	0.8	1.2

To make the portfolio gamma and vega neutral, both Option 1 and Option 2 can be used. If w_1 and w_2 are the quantities of Option 1 and Option 2 that are added to the portfolio, we require that

$$-5,000 + 0.5w_1 + 0.8w_2 = 0$$

$$-8,000 + 2.0w_1 + 1.2w_2 = 0$$

The solution to these equations is $w_1 = 400$, $w_2 = 6,000$. The portfolio can therefore be made gamma and vega neutral by including 400 of Option 1 and 6,000 of Option 2. The delta of the portfolio after the addition of the positions in the two traded options is $400 \times 0.6 + 6,000 \times 0.5 = 3,240$. Hence, 3,240 units of the underlying asset would have to be sold to maintain delta neutrality.

The vega of a long position in an option is positive. The variation of vega with the price of the underlying asset is similar to that of gamma and is shown in Figure 15.6. Gamma neutrality protects against large changes in the price of the underlying asset between hedge rebalancing. Vega neutrality protects against variations in volatility.

The implied volatilities of short-dated options tend to be more variable than the implied volatilities of long-dated options. The vega of a portfolio is therefore often calculated by changing the volatilities of short-dated options by more than that of long-dated options. One way of doing this is discussed in Section 8.10.

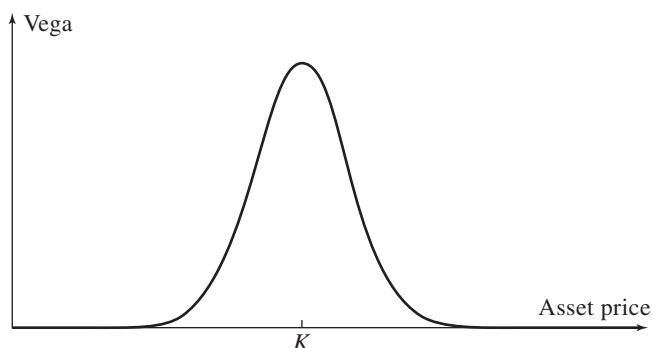


Figure 15.6 Variation of Vega of an Option with Price of Underlying Asset Where K Is Option's Strike Price

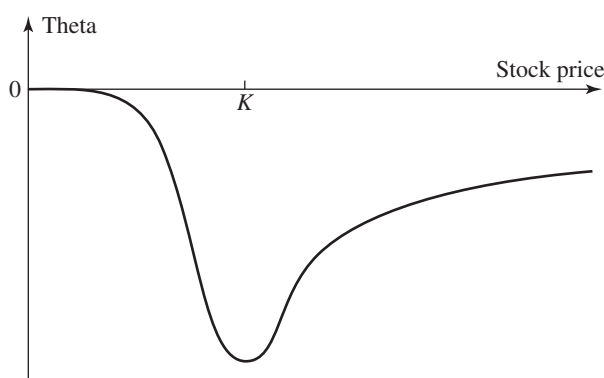


Figure 15.7 Variation of Theta of a European Call Option with Stock Price Where K Is Option's Strike Price

15.4 Theta

The *theta* of a portfolio, Θ , is the rate of change of the value of the portfolio with respect to the passage of time, with all else remaining the same. Theta is sometimes referred to as the *time decay* of the portfolio.

Theta is usually negative for an option.⁵ This is because as the time to maturity decreases with all else remaining the same, the option tends to become less valuable. The general way in which Θ varies with stock price for a call option on a stock is shown in Figure 15.7. When the stock price is very low, theta is close to zero. For an at-the-money call option, theta is large and negative. Figure 15.8 shows typical patterns for the variation of Θ with the time to maturity for in-the-money, at-the-money, and out-of-the-money call options.

⁵An exception to this could be an in-the-money European put option on a non-dividend-paying stock or an in-the-money European call option on a currency with a very high interest rate.

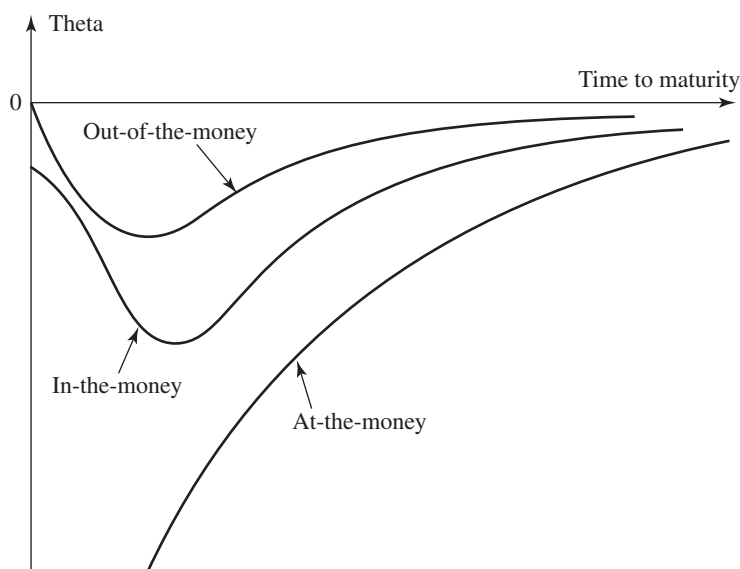


Figure 15.8 Typical Patterns for Variation of Theta of a European Call Option with Time to Maturity

Theta is not the same type of Greek letter as delta. There is uncertainty about the future price of the underlying asset, but there is no uncertainty about the passage of time. It makes sense to hedge against changes in the price of an underlying asset, but it does not make any sense to hedge against the effect of the passage of time on an option portfolio. In spite of this, many traders regard theta as a useful descriptive statistic for a portfolio. In a delta-neutral portfolio, when theta is large and positive, gamma tends to be large and negative, and vice versa.

15.5 Rho

Yet another Greek letter is rho. Rho is the rate of change of a portfolio with respect to the level of interest rates. Currency options have two rhos, one for the domestic interest rate and one for the foreign interest rate. When bonds and interest rate derivatives are part of the portfolio, traders usually consider carefully the ways in which the whole term structure of interest rates can change. We discussed this in chapter 14.

15.6 Calculating Greek Letters

Appendices E and F explain how Greek letters can be calculated for options. The RMFI software, which can be downloaded from the author's website, can be used for European and American options and some exotic options. Consider again the European call option in Section 15.1. The stock price is \$49, the strike price is \$50, the risk-free rate is 5%, the stock price volatility is 20%, and the time to exercise is 20 weeks or 20/52 year. Table 15.4

Table 15.4 Greek Letters Calculated Using RMFI Software

	Single Option	Short Position in 100,000 Options
Value (\$)	2.40	−240,000
Delta	0.522	−52,200
Gamma	0.066	−6,600
Vega (per %)	0.121	−12,100
Theta (per day)	−0.012	1,200
Rho (per %)	0.089	−8,900

shows delta, gamma, vega, theta, and rho for the option (i.e., for a long position in one option) and for a short position in 100,000 options, which was the position considered in Tables 15.2 and 15.3.

Here are some examples of how these numbers can be interpreted:

1. When there is an increase of \$0.1 in the stock price with no other changes, the option price increases by about 0.522×0.1 or \$0.0522. The value of a short position in 100,000 options decreases by \$5,220.
2. When there is an increase of \$0.1 in the stock price with no other changes, the delta of the option increases by about 0.066×0.1 or 0.0066. The delta of a short position in 100,000 options decreases by 660.
3. When there is an increase in volatility of 0.5% from 20% to 20.5% with no other changes, the option price increases by about 0.121×0.5 or \$0.0605. The value of a short position in 100,000 options decreases by \$6,050.
4. When one day goes by with no changes to the stock price or its volatility, the option price decreases by about \$0.012. The value of a short position in 100,000 options increases by \$1,200.
5. When interest rates increase by 1% (or 100 basis points) with no other changes, the option price increases by \$0.089. The value of a short position in 100,000 options decreases by \$8,900.

15.7 Taylor Series Expansions

Taylor series expansions are explained in Appendix G. They can be used to show how the change in the portfolio value in a short period of time depends on the Greek letters. Consider a portfolio dependent on a single asset price, S . If the volatility of the underlying asset and interest rates are assumed to be constant, the value of the portfolio, P , is a function of S , and time t . The Taylor series expansion gives

$$\Delta P = \frac{\partial P}{\partial S} \Delta S + \frac{\partial P}{\partial t} \Delta t + \frac{1}{2} \frac{\partial^2 P}{\partial S^2} \Delta S^2 + \frac{1}{2} \frac{\partial^2 P}{\partial t^2} \Delta t^2 + \frac{\partial^2 P}{\partial S \partial t} \Delta S \Delta t + \dots \quad (15.1)$$

where ΔP and ΔS are the change in P and S in a small time interval Δt . The first term on the right-hand side is delta times ΔS and is eliminated by delta hedging. The second

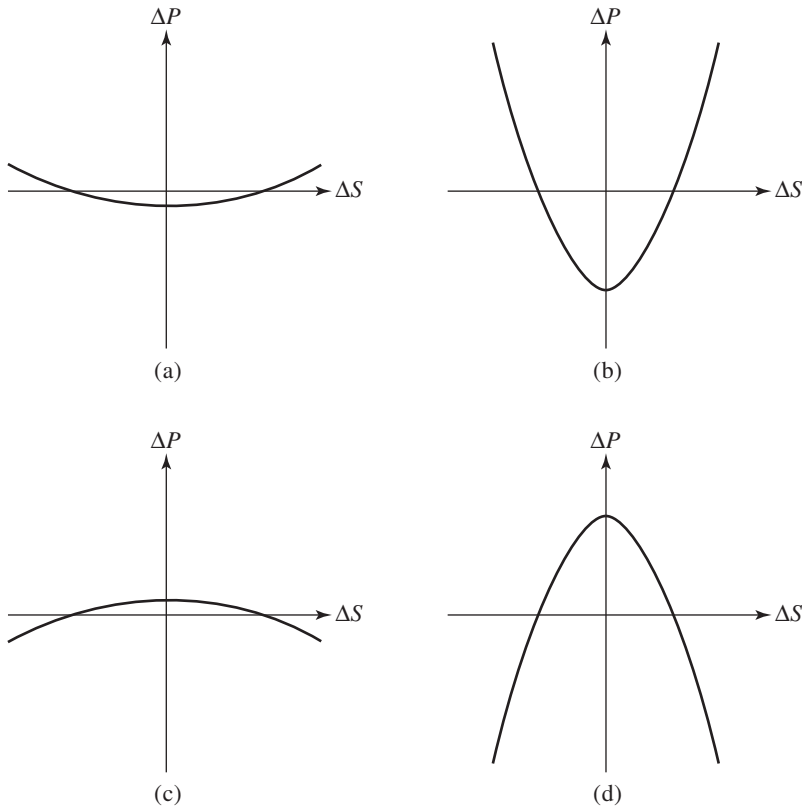


Figure 15.9 Alternative Relationships between ΔP and ΔS for a Delta-Neutral Portfolio (a) Slightly positive gamma, (b) large positive gamma, (c) slightly negative gamma, and (d) large negative gamma

term, which is theta times Δt , is non-stochastic. The third term can be made zero by ensuring that the portfolio is gamma neutral as well as delta neutral. Arguments from stochastic calculus show that ΔS is of order $\sqrt{\Delta t}$. This means that the third term on the right-hand side is of order Δt . Later terms in the Taylor series expansion are of higher order than Δt .

For a delta-neutral portfolio, the first term on the right-hand side of equation (15.1) is zero, so that

$$\Delta P = \Theta \Delta t + \frac{1}{2} \Gamma \Delta S^2 \quad (15.2)$$

when terms of higher order than Δt are ignored. The relationship between the change in the portfolio value and the change in the stock price is quadratic, as shown in Figure 15.9. When gamma is positive, the holder of the portfolio gains from large movements in the asset price and loses when there is little or no movement. When gamma is negative, the reverse is true so that a large positive or negative movement in the asset price leads to severe losses.

Example 15.2

Suppose that the gamma of a delta-neutral portfolio of options on an asset is $-10,000$. Suppose that a change of $+2$ in the price of the asset occurs over a short period of time (for which Δt can be assumed to be zero). Equation (15.2) shows that there is an unexpected decrease in the value of the portfolio of approximately $0.5 \times 10,000 \times 2^2 = \$20,000$. Note that the same unexpected decrease would occur if there were a change of -2 .

When the volatility, σ , of the underlying asset is uncertain, P is a function of σ , S , and t . Equation (15.1) then becomes

$$\Delta P = \frac{\partial P}{\partial S} \Delta S + \frac{\partial P}{\partial \sigma} \Delta \sigma + \frac{\partial P}{\partial t} \Delta t + \frac{1}{2} \frac{\partial^2 P}{\partial S^2} \Delta S^2 + \frac{1}{2} \frac{\partial^2 P}{\partial \sigma^2} \Delta \sigma^2 + \dots$$

where $\Delta \sigma$ is the change in σ in time Δt . In this case, delta hedging eliminates the first term on the right-hand side. The second term is eliminated by making the portfolio vega neutral. The third term is non-stochastic. The fourth term is eliminated by making the portfolio gamma neutral.

Traders often define other “Greek letters” to correspond to higher-order terms in the Taylor series expansion. For example, $\partial^2 P / \partial S \partial \sigma$ is known as *vanna*, $\partial^2 P / \partial \sigma^2$ is known as *vomma*, and $\partial^2 P / \partial S \partial t$ is known as *charm*.

15.8 The Realities of Hedging Derivatives

In an ideal world, traders working for financial institutions would be able to rebalance their portfolios very frequently in order to maintain a zero delta, a zero gamma, a zero vega, and so on. In practice, this is not possible. When managing a large portfolio dependent on a single underlying asset, traders usually make delta zero, or close to zero at least once a day by trading the underlying asset. Unfortunately a zero gamma and a zero vega are less easy to achieve because it is difficult to find options or other nonlinear derivatives that can be traded in the volume required at competitive prices (see discussion of dynamic hedging in Business Snapshot 15.2).

As noted earlier, there are big economies of scale in trading derivatives. Maintaining delta neutrality for an individual option on an asset by trading the asset daily would be prohibitively expensive. But it is realistic to do this for a portfolio of several hundred options on the asset. This is because the cost of daily rebalancing is covered by the profit on many different trades.

It is also worth noting that traders tend to work with implied volatilities (see Section 8.2) when calculating vega. When vega is calculated for a portfolio of options, a change in implied volatility is applied to each option. As already mentioned this change may be a decreasing function of option maturity.

BUSINESS SNAPSHOT 15.2

Dynamic Hedging in Practice

In a typical arrangement at a financial institution, the responsibility for a portfolio of derivatives dependent on a particular underlying asset is assigned to one trader or to a group of traders working together. For example, one trader at Goldman Sachs might be assigned responsibility for all derivatives dependent on the value of the Australian dollar. A computer system calculates the value of the portfolio and Greek letters for the portfolio. Limits are defined for each Greek letter and special permission is required if a trader wants to exceed a limit at the end of a trading day.

The delta limit is often expressed as the equivalent maximum position in the underlying asset. For example, the delta limit of Goldman Sachs on a stock might be specified as \$10 million. If the stock price is \$50, this means that the absolute value of delta as we have calculated it can be no more than 200,000. The vega limit is usually expressed as a maximum dollar exposure per 1% change in the volatility.

As a matter of course, options traders make themselves delta neutral—or close to delta neutral—at the end of each day. Gamma and vega are monitored, but are not usually managed on a daily basis. Financial institutions often find that their business with clients involves writing options and that as a result they accumulate negative gamma and vega. They are then always looking out for opportunities to manage their gamma and vega risks by buying options at competitive prices.

There is one aspect of an options portfolio that mitigates problems of managing gamma and vega somewhat. Options are often close to the money when they are first sold so that they have relatively high gammas and vegas. But after some time has elapsed, the underlying asset price has often changed enough for them to become deep-out-of-the-money or deep-in-the-money. Their gammas and vegas are then very small and of little consequence. The nightmare scenario for an options trader is where written options remain very close to the money as the maturity date is approached.

15.9 Hedging Exotic Options

Exotic options (see Section 5.5) can often be hedged using the approach we have outlined. As explained in Business Snapshot 15.3, delta hedging is sometimes easier for exotics and sometimes more difficult. When delta hedging is not feasible for a portfolio of exotic options an alternative approach known as *static options replication* is sometimes used. This is illustrated in Figure 15.10. Suppose that S denotes the asset price and t denotes time with the current ($t = 0$) value of S being S_0 . Static options

BUSINESS SNAPSHOT 15.3

Is Delta Hedging Easier or More Difficult for Exotics?

We can approach the hedging of exotic options by creating a delta-neutral position and rebalancing frequently to maintain delta neutrality. When we do this, we find that some exotic options are easier to hedge than plain vanilla options and some are more difficult.

An example of an exotic option that is relatively easy to hedge is an average price call option (see Asian options in Section 5.7). As time passes, we observe more of the asset prices that will be used in calculating the final average. This means that our uncertainty about the payoff decreases with the passage of time. As a result, the option becomes progressively easier to hedge. In the final few days, the delta of the option always approaches zero because price movements during this time have very little impact on the payoff.

By contrast, barrier options (see Section 5.7) are relatively difficult to hedge. Consider a knock-out call option on a currency when the exchange rate is 0.0005 above the barrier. If the barrier is hit, the option is worth nothing. If the barrier is not hit, the option may prove to be quite valuable. The delta of the option is discontinuous at the barrier, making conventional hedging quite difficult.

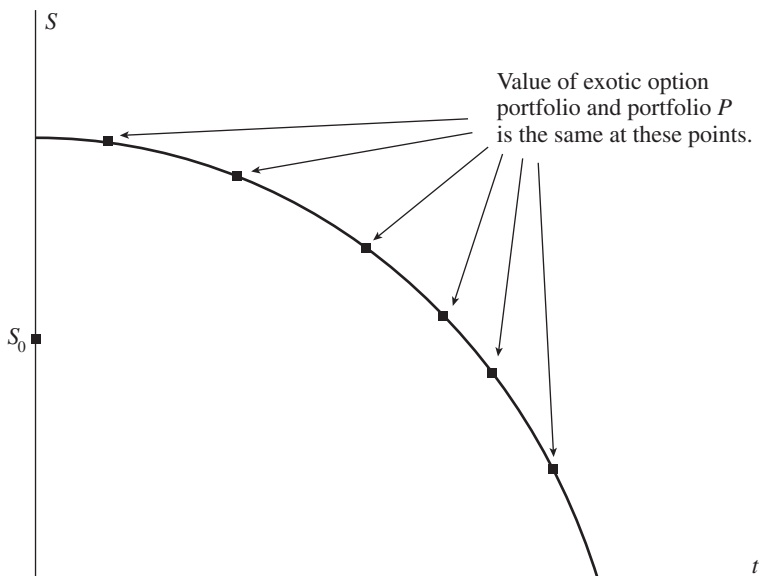


Figure 15.10 Static Options Replication

In a replicating portfolio, P is chosen so that it has the same value as the exotic option portfolio at a number of points on a boundary.

replication involves choosing a boundary in $\{S, t\}$ space that will eventually be reached and then finding a portfolio of regular options that is worth the same as the portfolio of exotic options at a number of points on the boundary. The portfolio of exotic options is hedged by shorting this portfolio of regular options. Once the boundary is reached, the hedge is unwound. A new hedge can then be created with static options replication if desired.

The theory underlying static options replication is that, if two portfolios are worth the same at all $\{S, t\}$ points on the boundary, they must be worth the same at all the $\{S, t\}$ points that can be reached prior to the boundary. In practice, values of the original portfolio of exotic options and the replicating portfolio of regular options are matched at some, but not all, points on the boundary. The procedure therefore relies on the idea that, if two portfolios have the same value at a reasonably large number of points on the boundary, their values are likely to be close at other points on the boundary.

15.10 Scenario Analysis

In addition to monitoring risks such as delta, gamma, and vega, option traders often also carry out a scenario analysis. The analysis involves calculating the gain or loss on their portfolio over a specified period under a variety of different scenarios. The time period chosen is likely to depend on the liquidity of the instruments. The scenarios can be either chosen by management or generated by a model.

Consider a trader with a portfolio of options on a particular foreign currency. There are two main variables on which the value of the portfolio depends. These are the exchange rate and the exchange rate volatility. Suppose that the exchange rate is currently 1.0000 and its volatility is 10% per annum. The bank could calculate a table such as Table 15.5 showing the profit or loss experienced during a two-week period under different scenarios. This table considers seven different exchange rates and three different volatilities.

In Table 15.5, the greatest loss is in the lower-right corner of the table. The loss corresponds to the volatility increasing to 12% and the exchange rate moving up to 1.06. Usually the greatest loss in a table such as 15.5 occurs at one of the corners, but

Table 15.5 Profit or Loss Realized in Two Weeks under Different Scenarios (\$ millions)

Volatility	Exchange Rate						
	0.94	0.96	0.98	1.00	1.02	1.04	1.06
8%	+102	+55	+25	+6	−10	−34	−80
10%	+80	+40	+17	+2	−14	−38	−85
12%	+60	+25	+9	−2	−18	−42	−90

this is not always so. For example, as we saw in Figure 15.9, when gamma is positive, the greatest loss is experienced when the underlying asset price stays where it is.

15.11 Approximate Analytical Results

The calculation of VaR and ES using the historical simulation approach when there are derivatives in a portfolio is straightforward. The main issue is that the valuation of derivatives for the various scenarios that are considered can be quite time consuming. As explained in Section 12.4, to speed up computations it is sometimes necessary to use a Taylor series expansion to approximate the change in the value of a derivative. (The Greek letters in such an expansion are generally known because analysts have calculated them for other risk management purposes.)

The model-building approach is difficult to use when there are derivatives in a portfolio. Consider a portfolio consisting of a single option on an asset with price S . A Taylor series expansion shows that it is approximately true that

$$\Delta P = \frac{\partial P}{\partial S} \Delta S + \frac{\partial P}{\partial \sigma} \Delta \sigma + \frac{1}{2} \frac{\partial^2 P}{\partial S^2} (\Delta S)^2 \quad (15.3)$$

(The theta term and terms corresponding to higher order derivatives are usually quite small.) Figure 15.11 shows the impact of the last term for a long call position. We suppose that the probability distribution for the price of the underlying asset at the end of one day is normal. The figure shows that probability distribution for the option price is positively

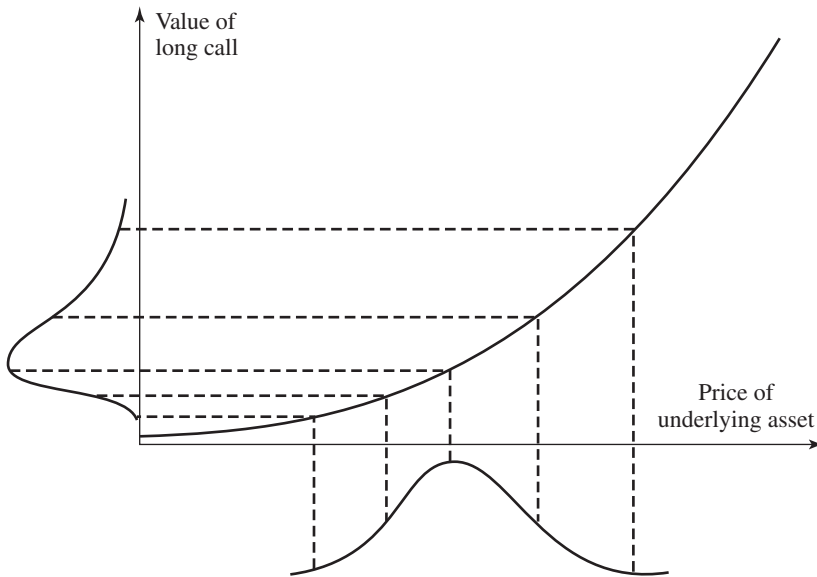


Figure 15.11 Translation of a Normal Probability Distribution for an Asset into a Probability Distribution for Value of a Long Call on the Asset

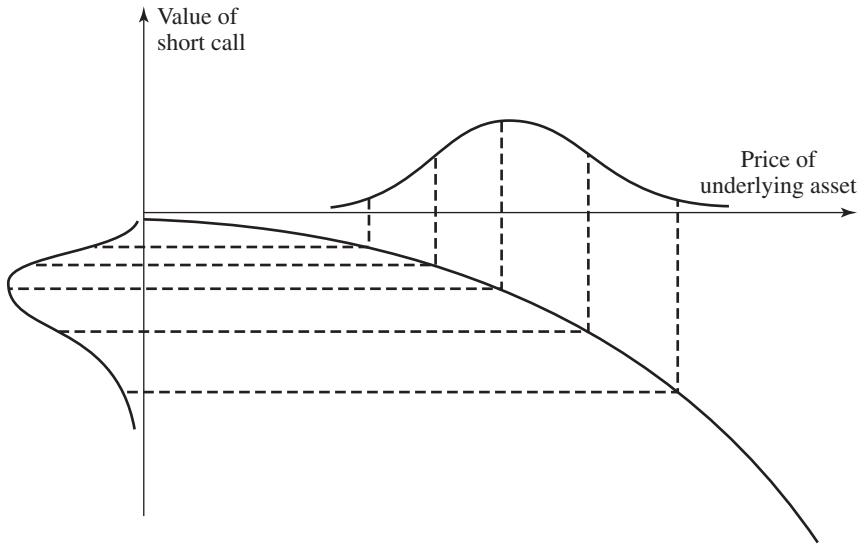


Figure 15.12 Translation of a Normal Probability Distribution for an Asset into a Probability Distribution for Value of a Short Call on the Asset

skewed. Figure 15.12 shows the relationship between the value of a short call position and the price of the underlying asset. A short call position has a negative gamma. In this case, we see that a normal distribution for the price of the underlying asset at the end of one day gets mapped into a negatively skewed distribution for the value of the option position.

The VaR for a portfolio is critically dependent on the left tail of the probability distribution of the portfolio value. For example, when the confidence level used is 99%, the VaR is the value in the left tail below which only 1% of the distribution resides. As indicated in Figure 15.11, a positive gamma portfolio tends to have a less heavy left tail than the normal distribution. If the distribution is assumed to be normal, the calculated VaR will tend to be too high. Similarly, as indicated in Figure 15.12, a negative gamma portfolio tends to have a heavier left tail than the normal distribution. If the distribution is assumed to be normal, the calculated VaR will tend to be too low.

15.11.1 Cornish–Fisher Expansions

A result in statistics known as the Cornish–Fisher expansion can be used to estimate percentiles of a probability distribution from its moments. As we will see, it can be useful for the determination of percentiles in non-linear models. Define μ_P and σ_P as the mean and standard deviation of ΔP so that

$$\mu_P = E(\Delta P)$$

$$\sigma_P^2 = E[(\Delta P)^2] - [E(\Delta P)]^2$$

The skewness of a probability distribution is related to the third moment. Positive skewness indicates that the right tail of the distribution is heavier than the left tail. Negative skewness indicates the reverse. The skewness of the probability distribution of ΔP , ξ_p , is usually defined as

$$\xi_p = \frac{1}{\sigma_p^3} E[(\Delta P - \mu_p)^3]$$

The skewness of a normal distribution is zero.

The kurtosis of a probability distribution is related to the fourth moment. It measures the heaviness of the tails of the distribution. Similarly to skewness, the kurtosis of the probability distribution of ΔP , κ_p , is usually defined as

$$\kappa_p = \frac{1}{\sigma_p^4} E[(\Delta P - \mu_p)^4]$$

The kurtosis of a normal distribution is 3. The *excess kurtosis* measures the kurtosis relative to the normal distribution and is defined as the kurtosis minus 3.

Using the first three moments of ΔP , the Cornish–Fisher expansion estimates the q -percentile of the distribution of ΔP as

$$\mu_p + w_q \sigma_p$$

where

$$w_q = z_q + \frac{1}{6}(z_q^2 - 1)\xi_p$$

and z_q is q -percentile of the standard normal distribution. Accuracy is improved when higher moments are used. For example, when kurtosis is considered, the expression for w_q becomes:

$$w_q = z_q + \frac{1}{6}(z_q^2 - 1)\xi_p + \frac{1}{24}(z_q^3 - 3z_q)(\kappa_p - 3)$$

Example 15.3

Suppose that for a certain portfolio we calculate $\mu_p = -0.2$, $\sigma_p = 2.2$, and $\xi_p = -0.4$, and we are interested in the one percentile ($q = 0.01$). In this case, $z_q = -2.326$. If we assume that the probability distribution of ΔP is normal, then the one percentile is

$$-0.2 - 2.326 \times 2.2 = -5.318$$

In other words, we are 99% certain that

$$\Delta P > -5.318$$

When we use the Cornish–Fisher expansion to adjust for skewness and set $q = 0.01$, we obtain

$$w_q = -2.326 - \frac{1}{6}(2.326^2 - 1) \times 0.4 = -2.620$$

so that the one percentile of the distribution is

$$-0.2 - 2.620 \times 2.2 = -5.965$$

Taking account of skewness, therefore, changes the VaR from 5.318 to 5.965. Suppose now that we also know that the kurtosis, κ_p , is 3.3. This changes our estimate of w_q to

$$w_q = -2.620 + \frac{1}{24}(-2.326^3 + 2.326) \times (3.3 - 3) = -2.691$$

so that the estimate of the one percentile becomes

$$-0.2 - 2.691 \times 2.2 = -6.119$$

15.11.2 Isserlis' Theorem

To apply the Cornish–Fisher result, we need to calculate moments of ΔP . When the changes are normal, Isserlis' theorem can in principle be used for this. The theorem states that if the X_i are zero-mean normal distributions

$$E(X_1 X_2 \cdots X_n) = 0$$

when n is odd and

$$E(X_1 X_2 \cdots X_n) = \sum \Pi E(X_i X_j)$$

when n is even. The notation $\sum \Pi E(X_i X_j)$ describes the result of (a) listing the distinct ways in which $X_1, X_2, \dots, X_n$ can be partitioned into $n/2$ pairs; (b) within each partition, calculating $E(X_i X_j)$ for each pair (X_i, X_j) and multiplying these expectations together; and (c) summing the results across partitions.

For example,

$$E(X_1 X_2 X_3 X_4) = E(X_1 X_2)E(X_3 X_4) + E(X_1 X_3)E(X_2 X_4) + E(X_1 X_4)E(X_2 X_3)$$

The expression for $E(X_1 X_2 X_3 X_4 X_5 X_6)$ has 15 terms, one of which is

$$E(X_1 X_2)E(X_3 X_4)E(X_5 X_6)$$

Isserlis' theorem is a potentially attractive result because we know the expectation of the product of a pair of zero-mean normally distributed variables, with standard deviations σ_X and σ_Y , is their covariance: $E(XY) = \rho \sigma_X \sigma_Y$ where ρ is the correlation between them.

Define Δx_i as the proportional change in the i th variable in time Δt . Define δ_i as the partial derivative of the portfolio value, P , with respect to the proportional change in the i th variable. (This corresponds to the definition in Equation (13.1) of Section 13.2.) Define γ_{ij} as the cross gamma calculated by first differentiating with respect to the proportional change in the i th variable and then differentiating with respect to the

proportional change in the j th variable. If σ_i is the standard deviation of Δx_i and ρ_{ij} is the correlation between Δx_i and Δx_j a Taylor Series expansion can be used to show that

$$E(\Delta P) = \frac{1}{2} \sum_{ij} \gamma_{ij} \rho_{ij} \sigma_i \sigma_j$$

$$E(\Delta P^2) = \sum_{ij} \delta_i \delta_j \sigma_i \sigma_j \rho_{ij} + \frac{1}{4} \sum_{ijkl} \gamma_{ij} \gamma_{kl} E(\Delta x_i \Delta x_j \Delta x_k \Delta x_l)$$

$$E(\Delta P^3) = \frac{3}{2} \sum_{ijkl} \delta_i \delta_j \gamma_{kl} E(\Delta x_i \Delta x_j \Delta x_k \Delta x_l) + \frac{1}{8} \sum_{ijklmn} \gamma_{ij} \gamma_{kl} \gamma_{mn} E(\Delta x_i \Delta x_j \Delta x_k \Delta x_l \Delta x_m \Delta x_n)$$

When there is only one risk factor so that we can dispense with the subscripts to σ , ρ , δ , and γ , using Isserlis' theorem these equations reduce to:

$$E(\Delta P) = \frac{1}{2} \gamma \sigma^2$$

$$E(\Delta P^2) = \delta^2 \sigma^2 + \frac{3}{4} \gamma^2 \sigma^4$$

$$E(\Delta P^3) = 4.5 \delta^2 \gamma \sigma^4 + 1.875 \gamma^3 \sigma^6$$

Unfortunately, the number of terms increases quickly as the number of risk factors increases, so that the calculation of the third moment becomes very time-consuming for even modest numbers of risk factors. However, some simplifications are achieved if cross gammas are set equal to zero so that $\gamma_{ij} = 0$ when $i \neq j$. The SIMM model, described in Appendix M, for determining initial margins in the bilaterally cleared OTC market makes use of these results.

Summary

A trader working for a bank, who is responsible for all the trades involving a particular asset, monitors a number of Greek letters and ensures that they are kept within the limits specified by the bank.

The delta, Δ , of a portfolio is the rate of change of its value with respect to the price of the underlying asset. Delta hedging involves creating a position with zero delta (sometimes referred to as a delta-neutral position). Because the delta of the underlying asset is 1.0, one way of hedging the portfolio is to take a position of $-\Delta$ in the underlying asset. For portfolios involving options and more complex derivatives, the position taken in the underlying asset has to be changed periodically. This is known as rebalancing.

Once a portfolio has been made delta neutral, the next stage is often to look at its gamma. The gamma of a portfolio is the rate of change of its delta with respect to the price of the underlying asset. It is a measure of the curvature of the relationship between the portfolio and the asset price. Another important hedge statistic is vega. This measures the rate of change of the value of the portfolio with respect to changes in the volatility of the underlying asset. Gamma and vega can be changed by trading options on the underlying asset.

In practice, derivatives traders usually rebalance their portfolios at least once a day to maintain delta neutrality. It is usually not feasible to maintain gamma and vega neutrality on a regular basis. Typically a trader monitors these measures. If they get too large, either corrective action is taken or trading is curtailed.

Further Reading

- Derman, E., D. Ergener, and I. Kani. "Static Options Replication." *Journal of Derivatives* 2, no. 4 (Summer 1995): 78–95.
- Passarelli, D. *Trading Option Greeks: How Time Volatility and Other Factors Drive Profits*. 2nd ed. Hoboken, NJ: John Wiley & Sons, 2012.
- Taleb, N. N. *Dynamic Hedging: Managing Vanilla and Exotic Options*. New York: John Wiley & Sons, 1997.

Practice Questions and Problems (Answers at End of Book)

- 15.1 The delta of a derivatives portfolio dependent on an index is $-2,100$. The index is currently 1,000. Estimate what happens to the value of the portfolio when the index increases to 1,005.
- 15.2 The vega of a derivatives portfolio dependent on the dollar–sterling exchange rate is 200 (per %). Estimate the effect on the portfolio of an increase in the volatility of the exchange rate from 12% to 14%.
- 15.3 The gamma of a delta-neutral portfolio is 30. Estimate what happens to the value of the portfolio when the price of the underlying asset (a) suddenly increases by \$2 and (b) suddenly decreases by \$2.
- 15.4 What does it mean to assert that the delta of a call option is 0.7? How can a short position in 1,000 options be made delta neutral when the delta of a long position in each option is 0.7?
- 15.5 What does it mean to assert that the theta of an option position is -100 per day? If a trader feels that neither a stock price nor its implied volatility will change, what type of option position is appropriate?
- 15.6 What is meant by the gamma of an option position? What are the risks in the situation where the gamma of a position is large and negative and the delta is zero?

- 15.7 “The procedure for creating an option position synthetically is the reverse of the procedure for hedging the option position.” Explain this statement.
- 15.8 A company uses delta hedging to hedge a portfolio of long positions in put and call options on a currency. Which of the following would lead to the most favorable result?
- (a) A virtually constant spot rate
 - (b) Wild movements in the spot rate
- How does your answer change if the portfolio contains short option positions?
- 15.9 A bank’s position in options on the dollar–euro exchange rate has a delta of 30,000 and a gamma of $-80,000$. Explain how these numbers can be interpreted. The exchange rate (dollars per euro) is 0.90. What position would you take to make the position delta neutral? After a short period of time, the exchange rate moves to 0.93. Estimate the new delta. What additional trade is necessary to keep the position delta neutral? Assuming the bank did set up a delta-neutral position originally, has it gained or lost money from the exchange-rate movement?
- 15.10 “Static options replication assumes that the volatility of the underlying asset will be constant.” Explain this statement.
- 15.11 Suppose that a trader using the static options replication technique wants to match the value of a portfolio of exotic derivatives with the value of a portfolio of regular options at 10 points on a boundary. How many regular options are likely to be needed? Explain your answer.
- 15.12 Why is an Asian option easier to hedge than a regular option?
- 15.13 Explain why there are economies of scale in hedging options.
- 15.14 Consider a six-month American put option on a foreign currency when the exchange rate (domestic currency per foreign currency) is 0.75, the strike price is 0.74, the domestic risk-free rate is 5%, the foreign risk-free rate is 3%, and the exchange-rate volatility is 14% per annum. Use the RMFI software (binomial tree with 100 steps) to calculate the price, delta, gamma, vega, theta, and rho of the option. Verify that delta is correct by changing the exchange rate to 0.751 and recomputing the option price.

Further Questions

- 15.15 The gamma and vega of a delta-neutral portfolio are 50 and 25, respectively, where the vega is “per %.” Estimate what happens to the value of the portfolio when there is a shock to the market causing the underlying asset price to decrease by \$3 and its volatility to increase by 4%.
- 15.16 Consider a one-year European call option on a stock when the stock price is \$30, the strike price is \$30, the risk-free rate is 5%, and the volatility is 25% per annum. Use the RMFI software to calculate the price, delta, gamma, vega, theta, and rho of the option. Verify that delta is correct by changing the stock price to \$30.1 and recomputing the option price. Verify that gamma is correct by recomputing the

delta for the situation where the stock price is \$30.1. Carry out similar calculations to verify that vega, theta, and rho are correct.

- 15.17 A financial institution has the following portfolio of over-the-counter options on pounds sterling:

Type	Position	Delta of Option	Gamma of Option	Vega of Option
Call	−1,000	0.50	2.2	1.8
Call	−500	0.80	0.6	0.2
Put	−2,000	−0.40	1.3	0.7
Call	−500	0.70	1.8	1.4

A traded option is available with a delta of 0.6, a gamma of 1.5, and a vega of 0.8.

- (a) What position in the traded option and in pounds sterling would make the portfolio both gamma neutral and delta neutral?
 - (b) What position in the traded option and in pounds sterling would make the portfolio both vega neutral and delta neutral?
- 15.18 Consider again the situation in Problem 15.17. Suppose that a second traded option, with a delta of 0.1, a gamma of 0.5, and a vega of 0.6, is available. How could the portfolio be made delta, gamma, and vega neutral?
- 15.19 Check the first three rows of Table 15.2 using the RMFI software. Calculate the gamma and theta of the position during the first three weeks, and the change in the value of the position (before the end-of-week rebalancing) during each of these weeks. Check whether equation (15.2) is approximately satisfied. (*Note:* The software produces a value of theta per calendar day. The theta in equation (15.2) is per year.)
- 15.20 A bank has a portfolio of options on an asset. The delta and gamma of the options (with respect to actual changes in the asset price) are -30 and -5 , respectively. Explain how these numbers can be interpreted. What are the delta and gamma with respect to proportional changes in the asset price? The asset price is 20 and its volatility is 1% per day. Using Isserlis' theorem, calculate the first three moments of the change in the daily portfolio value. Calculate a one-day 99% VaR (a) using the first two moments and (b) using the first three moments in conjunction with the Cornish–Fisher expansion.
- 15.21 Suppose that in Problem 15.20 the vega of the portfolio is -2 per 1% change in the annual volatility. Derive a model relating the change in the portfolio value in one day to delta, gamma, and vega.

Chapter 16

Scenario Analysis and Stress Testing

Stress testing involves evaluating the impact of extreme, but plausible, scenarios that are not considered by value at risk (VaR) or expected shortfall (ES) models. If there is one lesson to be learned from the market turmoil that started in the summer of 2007, it is that more emphasis should be placed on stress testing and less emphasis should be placed on the mechanistic application of VaR/ES models. VaR/ES models are useful, but they are inevitably backward looking. Risk management is concerned with what might happen in the future.

This chapter considers the different approaches that can be used to generate scenarios for stress testing and how the results should be used. It explains that the Global Financial Crisis of 2007–2008 has caused regulators to require banks to conduct more stress testing and that increasingly regulators are defining the stress tests themselves in an attempt to ensure that banks have sufficient capital to withstand adverse scenarios.

16.1 Generating the Scenarios

The most popular approach for calculating market risk VaR or ES is the historical simulation approach covered in Chapter 12. This approach assumes that data from the past are a good guide to what will happen over the next 1 to 10 days. But if an event has not occurred during the period covered by the data, it will not affect the results when the usual historical simulation method is used.

We have discussed a number of ways VaR/ES calculations can be modified so that they reflect more than the simple assumption that the future movements in market variables will be a random sample from the recent past. In particular:

1. Volatility scaling (see Section 12.3) can lead to more extreme outcomes being considered when the market is highly volatile.
2. Extreme value theory (see Section 12.5) provides a way of extending the tails of the loss distribution obtained from historical data.
3. Calculating stressed VaR or stressed ES (see Section 12.1) considers the impact of a particularly difficult period of 250 days that has occurred in the past.

But the nature of a VaR/ES calculation is that it is backward looking. Events that could happen, but are quite different from those that occurred during the period covered by the data, are not taken into account. Stress testing is an attempt to overcome this weakness of the VaR/ES measure.

Stress testing involves estimating how the portfolio of a financial institution would perform under scenarios involving extreme (but plausible) market moves. The exchange rate example in Table 8.1 illustrates the occurrence of extreme market moves. It shows five-standard-deviation moves happen about every five years on average, whereas they would happen hardly ever (about once every 7,000 years) if the distributions were normal.

A key issue in stress testing is the way in which scenarios are chosen. We now consider alternative procedures.

16.1.1 Stressing Individual Variables

One approach is to use scenarios where there is a large move in one variable and other variables are unchanged. Examples of scenarios of this type that are sometimes considered are:

1. A 100-basis-point parallel shift (up or down) in a yield curve.
2. Increasing or decreasing all the implied volatilities used for an asset by 50% of current values.
3. Increasing or decreasing an equity index by 10%.
4. Increasing or decreasing the exchange rate for a major currency by 6%.
5. Increasing or decreasing the exchange rate for a minor currency by 20%.

The impact of small changes in a variable is measured by its delta. The impact of larger changes can be measured by a combination of delta and gamma. Here we are considering changes that are so large that it is likely to be unreliable to estimate the change in the value of a portfolio using Greek letters.

16.1.2 Scenarios Involving Several Variables

Usually, when one market variable shows a big change, others do as well. This has led financial institutions to develop scenarios where several variables change at the same

time. A common practice is to use extreme movements in market variables that have occurred in the past. For example, to test the impact of an extreme movement in U.S. equity prices, a company might set the percentage changes in all market variables equal to those on October 19, 1987 (when the S&P 500 moved by 22.3 standard deviations). If this is considered to be too extreme, the company might choose January 8, 1988 (when the S&P 500 moved by 6.8 standard deviations). Other dates when there were big movements in equity prices are September 11, 2001, when terrorists attacked the World Trade Center in New York, and September 15, 2008, when Lehman Brothers declared bankruptcy. Equity prices also declined sharply in March 2020 because of uncertainties associated with the COVID-19 pandemic. The Dow Jones Industrial Average declined by 7.79%, 9.99%, and 12.93% on March 9, March 12, and March 16, respectively.

Another approach is to magnify what has happened in the past to generate extreme scenarios. For example, we might take a period of time when there were moderately adverse market movements and create a scenario where all variables move by three or five times as much as they did then. The problem with this approach is that correlations increase in stressed market conditions, and increasing the movements in all market variables by a particular multiple does not increase correlation.

Some historical scenarios are one-day shocks to market variables. Others involve shocks that take place over several days, several weeks, or even several months. It is important to include volatilities in the market variables that are considered. Typically, extreme movements in market variables such as interest rates and exchange rates are accompanied by large increases in the volatilities of these variables and large increases in the volatilities of a wide range of other variables. Some scenarios are likely to involve big movements in commodity prices, such as the plunge in oil prices in the second half of 2014. Others may involve a situation where there is a flight to quality combined with a shortage of liquidity and an increase in credit spreads. This was what happened in August 1998 when Russia defaulted on its debt and in July and August 2007 when investors lost confidence in the products created from the securitization of subprime mortgages (see Chapter 7).

16.1.3 Scenarios Generated by Management

History never repeats itself exactly. This may be partly because traders are aware of past financial crises and try to avoid making the same mistakes as their predecessors. The U.S. mortgage market led to the Global Financial Crisis that started in 2007. It is unlikely that future credit crises will be a result of mortgage-lending criteria being relaxed—but it is likely that there will be credit crises in the future.

In many ways, the scenarios that are most useful in stress testing are those generated by senior management or by the economics group within a financial institution. Senior management and the economics group are in a good position to use their understanding of markets, world politics, the economic environment, and current global uncertainties to develop plausible scenarios that would lead to large losses. Sometimes, the scenarios produced are based on things that have happened in the past, but are adjusted to include key features of the current financial and economic environment.

One way of developing the scenarios is for a committee of senior management to meet periodically and “brainstorm” answers to the simple question: “What can go wrong?” Clemens and Winkler (1999) have done studies to investigate the optimal composition of a committee of this type.¹ Their conclusions are that (a) the committee should have three to five members, (b) the backgrounds of the committee members should be different, and (c) there should be a healthy dialogue between members of the committee. It is important that members of the committee are able step back from their day-to-day responsibilities to consider the big picture.

It is not always the case that senior management’s thinking has to be highly innovative in order for it to come up with relevant scenarios. In 2005 and 2006, many commentators realized that the U.S. housing market was experiencing a bubble and that sooner or later the bubble would burst. It is easy to be wise after the event, but one reasonable scenario for the stress-testing committee to propose during that period would have been a 20% or 30% decline in house prices in all parts of the country.

It is important that senior management and the board of directors understand and recognize the importance of stress testing. They have the responsibility for taking strategic decisions based on the stress-testing results. One advantage of involving senior management in the development of the scenarios to be used in stress testing is that it should naturally lead to a “buy in” by them to the idea that stress testing is important. The results generated from scenarios that are created by individuals who have middle management positions are unlikely to be taken as seriously.

16.1.4 Core vs. Peripheral Variables

When individual variables are stressed or scenarios are generated by management, the scenarios are likely to be incomplete in that movements of only a few (core) market variables are specified. One approach is to set changes in all other (peripheral) variables to zero, but this is likely to be unsatisfactory. Another approach is to regress the peripheral variables on the core variables that are being stressed to obtain forecasts for them conditional on the changes being made to the core variables. These forecasts (as point forecasts or probability distributions) can be incorporated into the stress test.

This is known as *conditional stress testing* and is discussed by Kupiec (1999).² Kim and Finger (2000) carry this idea further by using what they call a “broken arrow” stress test. In this, the correlation between the core variables and the peripheral variables is based on what happens in stressed market conditions rather than what happens on average.³

16.1.5 Making Scenarios Complete

Scenarios should be carefully examined in an attempt to make sure that all the adverse consequences are considered. The scenarios should include not only the immediate

¹See R. Clemens and R. Winkler, “Combining Probability Distributions from Experts in Risk Analysis,” *Risk Analysis* 19, no. 2 (April 1999): 187–203.

²P. Kupiec, “Stress Testing in a Value at Risk Framework,” *Journal of Derivatives* 6 (1999): 7–24.

³See J. Kim and C. C. Finger, “A Stress Test to Incorporate Correlation Breakdown,” *Journal of Risk* 2, no. 3 (Spring 2000): 5–19.

BUSINESS SNAPSHOT 16.1**Long-Term Capital Management's Big Loss**

Long-Term Capital Management (LTCM), a hedge fund formed in the mid-1990s, experienced problems because of the way it provided collateral for its transactions. The hedge fund's investment strategy was known as convergence arbitrage. A very simple example of what it might do is the following. It would find two bonds, X and Y, issued by the same company promising the same payoffs, with X being less liquid (i.e., less actively traded) than Y. The market always places a value on liquidity. As a result, the price of X would be less than the price of Y. LTCM would buy X, short Y, and wait, expecting the prices of the two bonds to converge at some future time.

When interest rates increased, the company expected both bonds to move down in price by about the same amount so that the collateral it paid on bond X would be about the same as the collateral it received on bond Y. Similarly, when interest rates decreased, LTCM expected both bonds to move up in price by about the same amount so that the collateral it received on bond X would be about the same as the collateral it paid on bond Y. It therefore expected that there would be no significant outflow of funds as a result of its collateralization agreements.

In August 1998, Russia defaulted on its debt and this led to what is termed a "flight to quality" in capital markets. One result was that investors valued liquid instruments more highly than usual and the spreads between the prices of the liquid and illiquid instruments in LTCM's portfolio increased dramatically. The prices of the bonds LTCM had bought went down and the prices of those it had shorted increased. It was required to post collateral on both. The company was highly leveraged and found it difficult to make the payments required under the collateralization agreements. The result was that positions had to be closed out and there was a total loss of about \$4 billion. If the company had been less highly leveraged it would probably have been able to survive the flight to quality and could have waited for the prices of the liquid and illiquid bonds to become closer.

effect on the financial institution's portfolio of a shock to market variables, but also any "knock-on" effect resulting from many different financial institutions being affected by the shock in the same way and responding in the same way. Many people have said that they recognized the real estate bubble in the United States would burst in 2007, but did not realize how bad the consequences would be. They did not anticipate that many financial institutions would experience losses at the same time with the result that there

would be a flight to quality with severe liquidity problems and a huge increase in credit spreads.

Another example of knock-on effects is provided by the failure of Long-Term Capital Management (LTCM) in 1998 (see Business Snapshot 16.1). LTCM tended to have long positions in illiquid securities and short positions in liquid securities. Its failure was caused by a flight to quality following Russia's default on its debt. Many investors were only interested in buying liquid securities. Spreads between liquid and illiquid securities increased. LTCM had done stress tests looking at the impact of flights to quality similar to those that had occurred pre-1998. What it did not allow for was the knock-on effect. Many hedge funds were following similar trading strategies to those of LTCM in 1998. When the flight to quality started, they were all forced to unwind their positions at the same time. Unwinding meant selling illiquid securities and buying liquid securities, reinforcing the flight to quality and making it more severe than previous flights to quality.

Scenarios should ideally be dynamic so that the response to the scenario of the financial institution doing the stress test, as well as the responses of other financial institutions, are considered. Consider, for example, the situation where a financial institution has sold call options dependent on an underlying asset and maintains delta neutrality. A shock where there is a large increase in the asset price will lead to an immediate loss on the option position. To maintain delta neutrality, large amounts of the asset will have to be bought at a relatively high price. The cost of subsequent delta hedging is liable to depend on the volatility of the asset price. Often the volatility will be high, with big increases and decreases in the asset price before it settles down. The subsequent delta hedging is then particularly expensive.

16.1.6 Reverse Stress Testing

Reverse stress testing involves the use of computational procedures to search for scenarios that lead to a failure of the financial institution. It has become an important tool in risk management, and many bank supervisory bodies now require banks to have reverse stress testing programs in place.

Example 16.1

As a simple example of reverse stress testing, suppose a financial institution has positions in four European call options on an asset. The asset price is \$50, the risk-free rate is 3%, the volatility is 20%, and there is no income on the asset. The positions, strike prices, and lives of the options are as indicated in the table below. The current value of the position (in millions of dollars) is -25.90 . A numerical procedure can be used to search for one-day changes in the asset price and the volatility that will lead to the greatest losses. Some bounds should be put on the changes that are considered. We assume that

the asset price will not decrease below \$40 or increase above \$60. It is assumed that the volatility will not fall below 10% or rise above 30%.

Position (000s)	Strike Price	Life (years)	Position Value (\$ millions)
+250	50	1.0	1,176.67
−125	60	1.5	−293.56
−75	40	0.8	−843.72
−50	55	0.5	−65.30
Total			−25.90

Using the formulas for valuing European options (see Appendix E) in conjunction with Excel's Solver, the worst loss is found to be when the volatility decreases to 10% and the asset price falls to \$45.99. The loss is \$289.38 million. Reverse stress testing therefore shows that the financial institution is most exposed to a reduction of about 8% in the asset price combined with a sharp decline in volatility.

This might seem to be an unreasonable scenario. It is unlikely that volatility would go down sharply when the asset price declines by 8%. Solver could be run again with the lower bound to volatility being 20% instead of 10%. This gives a worst-case loss occurring when the volatility stays at 20% and the asset price falls to \$42.86. The loss is then \$87.19 million.

Searching over all the market variables to which a financial institution is exposed in the way indicated in Example 16.1 is in practice usually not computationally feasible. One approach is to identify 5 to 10 key market variables and assume that changes in other variables are dependent on changes in these variables.

Another way of simplifying the search process is to impose some structure on the problem. A principal components analysis (see Section 14.3) can be carried out on the changes in market variables (ideally using data from stressed market conditions) and then a search can be conducted to determine the changes in the principal components that generate large losses. This reduces the dimension of the space over which the search is conducted and should lead to fewer implausible scenarios.

An alternative approach is for the risk management group to impose a structure on scenarios. For example, a number of adverse scenarios that have occurred in the past can be specified. An analyst can then search to find what multiplier must be applied to the changes observed in all market variables under the scenarios in order for a particular loss level to be reached. For example, it might be concluded that a financial institution could survive a repeat of market movements during the Global Financial Crisis of 2008, but that if movements were 50% greater there would be serious problems.

Reverse stress testing can be used as a tool to facilitate brainstorming by the stress-testing committee. Prior to a meeting of the stress-testing committee, analysts can

use reverse stress testing to come up with several scenarios that would be disastrous to the financial institution. These scenarios, along with others they generate themselves, are then considered by the stress-testing committee. They use their judgment to eliminate some of the analysts' scenarios as implausible and modify others so that they become plausible and are retained for serious evaluation.

16.2 Regulation

We discuss the regulation of banks later in this book. At this stage it is worth noting that regulators require market risk calculations that are based on a bank's internal VaR models to be accompanied by "rigorous and comprehensive" stress testing. They make the point that stress testing is particularly important after long periods of benign conditions because such conditions tend to lead to complacency.

The recommendations of regulators highlight the importance of top management and board involvement in stress testing. In particular, top management and board members should be involved in setting stress-testing objectives, defining scenarios, discussing the results of stress tests, assessing potential actions, and decision making. Regulators make the point that the banks that fared well in the Global Financial Crisis that started in mid-2007 were the ones where senior management as a whole took an active interest in the development and operation of stress testing, with the results of stress testing serving as an input into strategic decision making. Stress testing should be conducted across all areas of the bank. It should not be the case that each area conducts its own stress test.

Specific recommendations for banks are:

1. Stress testing should form an integral part of the overall governance and risk management culture of the bank. Stress testing should be actionable, with the results from stress-testing analyses impacting decision making at the appropriate management level, including strategic business decisions of the board and senior management. Board and senior management involvement in the stress-testing program is essential for its effective operation.
2. A bank should operate a stress-testing program that promotes risk identification and control, provides a complementary risk perspective to other risk management tools, improves capital and liquidity management, and enhances internal and external communication.
3. Stress-testing programs should take account of views from across the organization and should cover a range of perspectives and techniques.
4. A bank should have written policies and procedures governing the stress-testing program. The operation of the program should be appropriately documented.
5. A bank should have a suitably robust infrastructure in place, which is sufficiently flexible to accommodate different and possibly changing stress tests at an appropriate level of granularity.

6. A bank should regularly maintain and update its stress-testing framework. The effectiveness of the stress-testing program, as well as the robustness of major individual components, should be assessed regularly and independently.
7. Stress tests should cover a range of risks and business areas, including at the firmwide level. A bank should be able to integrate effectively across the range of its stress-testing activities to deliver a complete picture of firmwide risk.
8. Stress-testing programs should cover a range of scenarios, including forward-looking scenarios, and aim to take into account system-wide interactions and feedback effects.
9. Stress tests should feature a range of severities, including events capable of generating most damage whether through size of loss or through loss of reputation. A stress-testing program should also determine what scenarios could challenge the viability of the bank (reverse stress tests) and thereby uncover hidden risks and interactions among risks.
10. As part of an overall stress-testing program, a bank should aim to take account of simultaneous pressures in funding and asset markets, and the impact of a reduction in market liquidity on exposure valuation.
11. The effectiveness of risk mitigation techniques should be systematically challenged.
12. The stress-testing program should explicitly cover complex and bespoke products such as securitized exposures. Stress tests for securitized assets should consider the underlying assets, their exposure to systematic market factors, relevant contractual arrangements and embedded triggers, and the impact of leverage, particularly as it relates to the subordination level in the issue structure.
13. The stress-testing program should cover pipeline and warehousing risks.⁴ A bank should include such exposures in its stress tests regardless of their probability of being securitized.
14. A bank should enhance its stress-testing methodologies to capture the effect of reputational risk. The bank should integrate risks arising from off-balance-sheet vehicles and other related entities in its stress-testing program.
15. A bank should enhance its stress-testing approaches for highly leveraged counterparties in considering its vulnerability to specific asset categories or market movements and in assessing potential wrong-way risk related to risk-mitigating techniques.

The recommendations for bank supervisors are:

16. Supervisors should make regular and comprehensive assessments of a bank's stress-testing programs.
17. Supervisors should require management to take corrective action if material deficiencies in the stress-testing program are identified or if the results of stress tests are not adequately taken into consideration in the decision-making process.

⁴“Pipeline and warehousing” risks refer to risks associated with assets that are awaiting securitization, but might not be securitized if market conditions change. These risks led to losses during the onset of the crisis.

18. Supervisors should assess and, if necessary, challenge the scope and severity of firm-wide scenarios. Supervisors may ask banks to perform sensitivity analysis with respect to specific portfolios or parameters, use specific scenarios, or evaluate scenarios under which their viability is threatened (reverse stress-testing scenarios).
19. Under the Pillar 2 (supervisory review process) of the Basel II framework, supervisors should examine a bank's stress-testing results as part of a supervisory review of both the bank's internal capital assessment and its liquidity risk management. In particular, supervisors should consider the results of forward-looking stress testing for assessing the adequacy of capital and liquidity.
20. Supervisors should consider implementing stress-test exercises based on common scenarios.
21. Supervisors should engage in a constructive dialogue with other public authorities and the industry to identify systemic vulnerabilities. Supervisors should also ensure that they have the capacity and the skills to assess banks' stress-testing programs.

16.2.1 Scenarios Chosen by Regulators

Bank regulators require banks to consider extreme scenarios and then make sure they have enough capital for those scenarios. There is an obvious problem here. Banks want to keep their regulatory capital as low as possible. They therefore have no incentive to consider extreme scenarios that would lead to a bank supervisor telling them that their capital requirements need to be increased. There is therefore a natural tendency for the scenarios they consider to be “watered down” and fairly benign.

One approach to overcoming this problem is for regulators themselves to provide the scenarios (see Recommendations 18 and 20). Regulators in many jurisdictions (including the United Kingdom, the European Union, and the United States) now regularly carry out their own stress tests to determine whether the banks they are supervising have enough capital to survive extreme adverse events. If banks fail the stress test, they are required to raise more capital and may be subject to dividend restrictions.

In the United States, the Federal Reserve has since 2009 carried out an annual stress test for all banks with consolidated assets of more than \$50 billion (i.e., G-SIBs and D-SIBs). Since 2011 this has been called the Comprehensive Capital Analysis and Review (CCAR). Banks are required to submit a capital plan (including planned dividends). The scenarios created by the Federal Reserve are recession scenarios similar to those in 1973–1975, 1981–1982, and 2007–2009. The scenarios include projections for about 25 variables, including macroeconomic variables such as gross domestic product (GDP) growth, unemployment rate, stock market indices, and house price indices.

Banks in the United States with consolidated assets over \$10 billion are subject to the Dodd–Frank Act Stress Test (DFAST). The scenarios considered in DFAST are similar to those in CCAR. However, banks do not have to submit a capital plan, as the management of capital is assumed to be based on a standard set of assumptions. Both CCAR and DFAST involve a huge amount of work for banks with thousands of pages of justification for the models used being produced.

BUSINESS SNAPSHOT 16.2**Traffic Light Options**

In June 2001, the Danish Financial Supervisory Authority (DFSA) introduced a “traffic light” solvency stress-testing system. This requires life insurance companies and pension funds to submit semiannual reports indicating the impact on them of certain predefined shocks. The “red light scenario” involves a 70-basis-point decrease in interest rates, a 12% decline in stock prices, and an 8% decline in real estate prices. If capital falls below a specified critical level in this scenario, the company is categorized with “red light status” and is subject to more frequent monitoring with monthly reports being required. The “yellow light scenario” involves a 100-basis-point decrease in interest rates, a 30% decline in stock prices, and a 12% decline in real estate prices. If capital falls below the critical level in this scenario, the company is categorized with “yellow light status” and has to submit quarterly reports. When the company’s capital stays above the critical levels for the red and yellow light scenarios, the company has a “green light status” and is subject to normal semiannual reporting.

Derivatives dealers have developed products for helping life insurance and pension funds keep a green light status. These are known as traffic light options. They pay off in the traffic light scenarios so as to give a boost to the financial institution’s performance when these scenarios are considered. Traffic light options pay off only when the movements close to those specified by regulators in the red light or yellow light scenario occur. The options are therefore not too expensive.

By choosing scenarios themselves, regulators are able to focus the attention of banks on recession scenarios that are of concern to regulators. If regulators see many banks taking positions with similar risks, they can require all banks to consider a particular set of scenarios that give rise to adverse results for the positions. The downside of regulators generating scenarios themselves is that part of the reason for the increased focus by supervisors on stress testing is that they want to encourage financial institutions to spend more time generating and worrying about potential adverse scenarios. If supervisors do the work in generating the scenarios, this may not happen and banks will focus only on the scenarios generated by regulators. In an ideal regulatory environment, both management-generated scenarios and supervisor-generated scenarios are evaluated.

There is a danger that financial institutions will find actions they can take that allow them to pass the stress tests without making them any safer. An extreme example of

this is in Business Snapshot 16.2. When regulators defined key scenarios for life insurance companies and pension funds in Denmark, some of those companies responded by hedging against the particular scenarios used by regulators—and only against those scenarios.⁵ This is not what regulators intended. Each scenario used in stress testing should be viewed as representative of a range of things that might happen. Financial institutions should ensure that their capital will be in good shape not just for the specified scenarios, but also for other similar or related scenarios. Regulators are now well aware of this type of problem and have procedures aimed at ensuring that banks do not game the system in this way.

16.3 What to Do with the Results

The biggest problem in stress testing is using the results effectively. All too often, the results of stress testing are ignored by senior management. A typical response is, “Yes, there are always one or two scenarios that will sink us. We cannot protect ourselves against everything that might happen.” One way of trying to avoid this sort of response is to involve senior management in the development of scenarios, as outlined earlier. A better response on the part of senior management would be, “Are the risks associated with these scenarios acceptable? If not, let’s investigate what trades we can put on or other actions we can take to make these types of risks more acceptable.”

The problem for both senior management and the risk management group is that they have two separate reports on their desks concerning what could go wrong. One report comes from VaR or ES models, the other from internal stress testing. Which one should they base their decision making on?

16.3.1 Integrating Stress Testing and VaR Calculations

Berkowitz (2000) suggests that stress testing will be taken more seriously if its results are integrated into the calculation of VaR.⁶ This can be done by assigning a probability to each stress scenarios that is considered. Suppose that a financial institution has considered n_s stress scenarios and the total probability assigned to the stress scenarios is p . Assume further that there are n_v VaR scenarios generated using historical simulation in the usual way. An analyst can assume that there are a total of $n_s + n_v$ scenarios. The n_s stress scenarios have probability p and the n_v historical scenarios have probability $1 - p$.

⁵The information in Business Snapshot 16.2 is from P. L. Jorgensen, “Traffic Light Options,” *Journal of Banking and Finance* 31, no. 12 (December 2007): 3698–3719.

⁶See J. Berkowitz, “A Coherent Framework for Stress Testing,” *Journal of Risk* 2, no. 2 (Winter 1999/2000): 5–15.

Unfortunately human beings are not good at estimating a subjective probability for the occurrence of a rare event. To make the task feasible for the stress-testing committee, one approach is to ask the stress-testing committee to allocate each scenario to categories with preassigned probabilities. The categories might be:

1. Probability = 0.05%. Extremely unlikely. One chance in 2,000.
2. Probability = 0.2%. Very unlikely, but the scenario should be given about the same weight as the 500 scenarios used in the historical simulation analysis.
3. Probability = 0.5%. Unlikely, but the scenario should be given more weight than the 500 scenarios used in the historical simulation analysis.

Example 16.2

Suppose that, in the example in Section 12.1, five stress scenarios are considered. They lead to losses (\$000s) of 400, 700, 900, 1,050, and 1,200. The probabilities assigned to the scenarios are 0.5%, 0.2%, 0.2%, 0.05%, and 0.05%, respectively. The total probability of the stress scenarios is, therefore, 1%. This means that the probability assigned to the scenarios generated by historical simulation is 99%. Assuming that equal weighting is used, each historical simulation scenario is assigned a probability of $0.99/500 = 0.00198$. Table 12.4 is therefore replaced by Table 16.1. The probabilities assigned to scenarios are

Table 16.1 Losses Ranked from Highest to Lowest

Scenario	Loss (\$000s)	Probability	Cumulative Probability
s5	1,200.000	0.00050	0.00050
s4	1,050.000	0.00050	0.00100
v427	922.484	0.00198	0.00298
s3	900.000	0.00200	0.00498
v429	858.423	0.00198	0.00696
s2	700.000	0.00200	0.00896
v424	653.542	0.00198	0.01094
v415	490.215	0.00198	0.01292
v482	422.291	0.00198	0.01490
s1	400.000	0.00500	0.01990
v440	362.733	0.00198	0.02188
v426	360.532	0.00198	0.02386
v431	353.788	0.00198	0.02584
...	...	...	...
...	...	...	...
...	...	...	...

For Example 16.2, s1, s2, ... are the stress scenarios; v1, v2, ... are the VaR historical simulation scenarios.

accumulated from the worst scenario to the best.⁷ The VaR level when the confidence level is 99% is the first loss for which the cumulative probability is greater than 0.01. In our example this is \$653,542.

Rebonato (2010) suggests a more elaborate approach to assessing probabilities of scenarios involving a careful application of a well-known result in statistics, Bayes' theorem, and what are known as Bayesian networks.⁸ The probability of a scenario consisting of two events is equal to the probability of the first event happening times the probability of the second event happening conditional on the first event having happened. Similarly the probability of a scenario consisting of three events is the probability of the first event happening times the probability of the second event happening conditional that the first event has happened times the probability of the third event happening conditional that the first two events have happened. Rebonato's approach provides a way of evaluating the conditional probabilities.

16.3.2 *Subjective vs. Objective Probabilities*

There are two types of probability estimates: objective and subjective. An *objective probability* for an event is a probability calculated by observing the frequency with which the event happens in repeated trials. As an idealized example of an objective probability, consider an urn containing red balls and black balls in an unknown proportion. We want to know the probability of a ball drawn at random from the urn being red. We could draw a ball at random, observe its color, replace it in the urn, draw another ball at random, observe its color, replace it in the urn, and so on, until 100 balls have been drawn. If 30 of the balls that have been drawn are red and 70 are black, our estimate for the probability of drawing a red ball is 0.3. Unfortunately, most objective probabilities calculated in real life are usually less reliable than the probability in this example, because the probability of the event happening does not remain constant for the observations that are available and the observations may not be independent.

A *subjective probability* is a probability derived from an individual's personal judgment about the chance of a particular event occurring. The probability is not based on historical data. It is a degree of belief. Different people are liable to have different subjective probabilities for the same event.

The probabilities in historical simulation are objective whereas the probabilities assigned to the scenarios in stress testing are subjective. Many analysts are uncomfortable with subjective probabilities because they are not based on data. Also, it is unfortunately the case that political considerations may play a part in a financial institution's decision to focus on historical data. If you use historical data and things go wrong, you can

⁷This is the same procedure that we used when weights were assigned to historical simulation scenarios. (See Table 13.5 in Section 13.3.)

⁸See Riccardo Rebonato, *Coherent Stress Testing: A Bayesian Approach to Financial Stress* (Chichester, UK: John Wiley & Sons, 2010).

blame the data. If you use subjective judgments that have been provided by a group of people, those people are liable to be blamed.

However, if it is based only on objective probabilities, risk management is inevitably backward looking and fails to capitalize on the judgment and expertise of senior managers. It is the responsibility of those managers to steer the financial institution so that catastrophic risks are avoided.

Summary

Stress testing is an important part of the risk management process. It leads to a financial institution considering the impact of extreme scenarios that are ignored by a traditional VaR or ES analysis, but that do happen from time to time. Once plausible scenarios have been evaluated, the financial institution can take steps to lessen the impact of the particularly bad ones. One advantage of a comprehensive stress-testing program is that a financial institution obtains a better understanding of the nature of the risks in its portfolio.

Scenarios can be generated in a number of different ways. One approach is to consider extreme movements in just one market variable while keeping others fixed. Another is to use the movements in all market variables that occurred during periods in the past when the market experienced extreme shocks. The best approach is to ask a committee of senior management and economists to use their judgment and experience to generate the plausible extreme scenarios. Sometimes financial institutions carry out reverse stress testing where algorithms are used to search for scenarios that would lead to large losses. Scenarios should be as complete as possible and include the impact of knock-on effects as well as the initial shock to market variables. The market turmoil starting in summer 2007 shows that, in some cases, the knock-on effect can be significant and include a flight to quality, an increase in credit spreads, and a shortage of liquidity.

Regulators require financial institutions to keep sufficient capital for stress scenarios. Regulators in the United States, European Union, and United Kingdom regularly develop stress scenarios to be evaluated by all the large financial institutions they are responsible for. This may lead to some of the financial institutions having to raise more capital and has the potential to uncover systemic risk problems.

If subjective probabilities are assigned to the extreme scenarios that are considered, stress testing can be integrated with a VaR analysis. This is an interesting idea, but it does not seem to have been widely adopted by financial institutions or their regulators.

Further Reading

- Alexander, C., and E. A. Sheedy. "Developing a Stress-Testing Framework Based on Market Risk Models." *Journal of Banking and Finance* 32, no. 10 (2008): 2220–2236.
- Aragónés, J. R., C. Blanco, and K. Dowd. "Incorporating Stress Tests into Market Risk Modeling." *Derivatives Quarterly* 7 (Spring 2001): 44–49.

- Aragónés, J. R., C. Blanco, and K. Dowd. "Stress Tests, Market Risk Measures, and Extremes: Bringing Stress Tests to the Forefront of Market Risk Management." In *Stress Testing for Financial Institutions: Applications, Regulations, and Techniques*, edited by D. Rösch and H. Scheule. London: Risk Books, 2008.
- Basel Committee on Banking Supervision. "Principles for Sound Stress-Testing Practices and Supervision," May 2009.
- Berkowitz, J. "A Coherent Framework for Stress Testing." *Journal of Risk* 2, no. 2 (Winter 1999/2000): 5–15.
- Bogle, J. C. "Black Monday and Black Swans." *Financial Analysts Journal* 64, no. 2 (March/April 2008): 30–40.
- Clemens, R., and R. Winkler. "Combining Probability Distributions from Experts in Risk Analysis." *Risk Analysis* 19, no. 2 (April 1999): 187–203.
- Duffie, D. "Systemic Risk Exposures: A 10-by-10-by-10 Approach." Working Paper, Stanford University, 2011.
- Glasserman, P., C. Kang, and W. Kang. "Stress Scenario Selection by Empirical Likelihood." *Quantitative Finance* 15 (2015): 25–41.
- Hassani, B. *Scenario Analysis in Risk Management*. Cham, Switzerland: Springer, 2016.
- Hua, P., and P. Wilmott. "Crash Courses." *Risk* 10, no. 6 (June 1997): 64–67.
- Kim, J., and C. C. Finger. "A Stress Test to Incorporate Correlation Breakdown." *Journal of Risk* 2, no. 3 (Spring 2000): 5–19.
- Kupiec, P. "Stress Testing in a Value at Risk Framework." *Journal of Derivatives* 6 (1999): 7–24.
- Rebonato, R. *Coherent Stress Testing: A Bayesian Approach to the Analysis of Financial Stress*. Chichester, UK: John Wiley & Sons, 2010.
- Taleb, N. N. *The Black Swan: The Impact of the Highly Improbable*. New York: Random House, 2007.

Practice Questions and Problems (Answers at End of Book)

- 16.1 Explain three different ways that scenarios can be generated for stress testing.
- 16.2 What is reverse stress testing? How is it used?
- 16.3 Why might the regulatory environment lead to a financial institution underestimating the severity of the scenarios it considers?
- 16.4 What are traffic light options? What are their drawbacks?
- 16.5 Why is it important for senior management to be involved in stress testing?
- 16.6 What are the advantages and disadvantages of bank regulators choosing some of the scenarios that are considered for stress testing?
- 16.7 Explain the difference between subjective and objective probabilities.
- 16.8 Suppose that, in the example in Section 12.1, seven stress scenarios are considered. They lead to losses (\$000s) of 400, 600, 700, 1,000, 1,100, 1,300, and 1,400. The subjective probabilities assigned to the scenarios are 0.5%, 0.5%, 0.2%, 0.2%, 0.05%, 0.05%, and 0.05%, respectively. What is the new one-day 99% VaR that would be calculated using the procedure discussed in Section 16.3?
- 16.9 Suppose that the positions in the four options in Example 16.1 are changed to 200, –70, –120, and –60. Use Solver to calculate the worst-case scenario for a daily change. Consider asset prices between \$40 and \$60 and volatilities between 10% and 30%.

Further Questions

- 16.10 What difference does it make to the worst-case scenario in Example 16.1 if (a) the options are American rather than European and (b) the options are barrier options that are knocked out if the asset price reaches \$65? Use Solver to search over asset prices between \$40 and \$60 and volatilities between 18% and 30%.
- 16.11 What difference does it make to the VaR calculated in Example 16.2 if the exponentially weighted moving average model is used to assign weights to historical scenarios as described in Section 12.3.1?

Part Four

CREDIT RISK

Chapter 17

Estimating Default Probabilities

In this chapter, we discuss a number of different approaches to estimating default probabilities and explain the key difference between risk-neutral and real-world estimates. The material we cover will be used in Chapter 18 when we consider the counterparty credit risk in derivative transactions, and in Chapter 19 when we discuss the calculation of credit value at risk.

17.1 Credit Ratings

As explained in Section 1.7, rating agencies such as Moody's, S&P, and Fitch provide ratings describing the creditworthiness of corporate bonds.¹ Because a credit rating is designed to provide information about credit quality, one might expect frequent changes in credit ratings as positive and negative information reaches the market. In fact, ratings change relatively infrequently. One of the objectives of rating agencies when they assign ratings is stability. For example, they want to avoid ratings reversals, where a company's bonds are downgraded and then upgraded a few weeks later. Ratings therefore change only when there is reason to believe that a long-term change in the company's

¹In theory, a credit rating is an attribute of a bond issue, not a company. However, in many cases all bonds issued by a company have the same rating. A rating is therefore often referred to as an attribute of a company.

creditworthiness has taken place. The reason for this is that bond traders are major users of ratings. Often they are subject to rules governing what the credit ratings of the bonds they hold must be. (For example, many bond funds are allowed to hold only investment-grade bonds.) If these ratings changed frequently, they might have to do a large amount of trading (and incur high transaction costs) just to satisfy the rules.

A related point is that rating agencies try to “rate through the cycle.” Suppose that the economy exhibits a downturn and this has the effect of increasing the probability of a company defaulting in the next six months, but makes very little difference to the company’s probability of defaulting over the next three to five years. A rating agency would not usually change the company’s credit rating in these circumstances.

Companies such as Moody’s KMV and Kamakura provide estimates of a company’s probability of default that are based on its equity price and other variables. These estimates do not have stability as one of their objectives and tend to respond more quickly to market information than credit ratings. The types of models that are used to produce the estimates will be discussed in Section 17.8.

17.1.1 *Internal Credit Ratings*

Most banks have procedures for rating the creditworthiness of their corporate and retail clients. This is a necessity. The ratings published by rating agencies are usually available only for companies that have issued publicly traded debt. As a result, many small and medium-sized companies do not have credit ratings from rating agencies.

Internal-ratings-based approaches for estimating probabilities of default typically involve profitability ratios such as return on assets and balance sheet ratios such as current assets divided by current liabilities (the current ratio) and debt to equity. Banks recognize that it is cash rather than profits that is necessary to repay a loan. They typically take the financial information provided by a company and convert it to a cash flow statement. This allows them to estimate how easy it will be for a company to service its debt.

17.1.2 *Altman’s Z-Score*

Edward Altman has pioneered the use of accounting ratios to predict default. In 1968, he developed what has become known as the Z-score.² Using a statistical technique known as discriminant analysis, he attempted to predict defaults from five accounting ratios:

X_1 : Working capital/Total assets

X_2 : Retained earnings/Total assets

X_3 : Earnings before interest and taxes/Total assets

X_4 : Market value of equity/Book value of total liabilities

X_5 : Sales/Total assets

²See E. I. Altman, “Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy,” *Journal of Finance* 23, no. 4 (September 1968): 589–609.

For publicly traded manufacturing companies, the original Z-score was

$$Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 0.999X_5 \quad (17.1)$$

If the Z-score was greater than 3.0, the company was considered unlikely to default. If it was between 2.7 and 3.0, there was reason to be “on alert.” If it was between 1.8 and 2.7, there was a good chance of default. If it was less than 1.8, the probability of a financial embarrassment was considered to be very high. Equation (17.1) was estimated from a sample of 66 publicly traded manufacturing companies. Of these, 33 failed within one year and 33 did not fail within one year. The model proved to be fairly accurate when tested out of sample (i.e., on a set of firms different from that used to estimate equation 17.1). Both Type I errors (companies that were predicted not to go bankrupt but did do so) and Type II errors (companies that were predicted to go bankrupt, but did not do so) were small.³ Variations on the model have been developed for manufacturing companies that are not publicly traded and for non-manufacturing companies. The Z-score methodology has been revised and extended since Altman’s original research more than 50 years ago.

Example 17.1

Consider a company for which working capital is 170,000, total assets are 670,000, earnings before interest and taxes is 60,000, sales are 2,200,000, the market value of equity is 380,000, total liabilities is 240,000, and retained earnings is 300,000. In this case, $X_1 = 0.254$, $X_2 = 0.448$, $X_3 = 0.0896$, $X_4 = 1.583$, and $X_5 = 3.284$. The Z-score is

$$1.2 \times 0.254 + 1.4 \times 0.448 + 3.3 \times 0.0896 + 0.6 \times 1.583 + 0.999 \times 3.284 = 5.46$$

The Z-score indicates that the company is not in danger of defaulting in the near future.

17.2 Historical Default Probabilities

Table 17.1 is typical of the data that are produced by rating agencies. It shows the default experience through time of companies that started with a certain credit rating. For example, Table 17.1 shows that a bond with an initial S&P credit rating of BBB has a 0.16% chance of defaulting by the end of the first year, a 0.43% chance of defaulting by the end of the second year, and so on. The probability of a bond defaulting during a particular year can be calculated from the table. For example, the probability that a bond initially rated BBB will default during the second year of its life is $0.43 - 0.16 = 0.27\%$.

Table 17.1 shows that, for investment-grade bonds (i.e., those rated BBB and above), the probability of default in a year tends to be an increasing function of time during the first few years. (For example, the probabilities of a bond initially rated A defaulting during

³Type I errors are liable to be more costly to the lending department of a commercial bank than Type II errors.

Table 17.1 Average Cumulative Issuer-Weighted Default Rates (%), 1981–2020

Rating	Time Horizon (years)							
	1	2	3	4	5	7	10	15
AAA	0.00	0.03	0.13	0.24	0.34	0.51	0.70	0.90
AA	0.02	0.06	0.11	0.21	0.30	0.49	0.70	0.99
A	0.05	0.13	0.22	0.33	0.46	0.76	1.20	1.84
BBB	0.16	0.43	0.75	1.14	1.54	2.27	3.24	4.54
BB	0.63	1.93	3.46	4.99	6.43	8.89	11.64	14.65
B	3.34	7.80	11.75	14.89	17.35	20.99	24.62	28.24
CCC/C	28.30	38.33	43.42	46.36	48.58	50.75	52.76	54.76

SOURCE: S&P Global Ratings Research.

years one, two, three, and four are 0.05%, 0.08%, 0.09%, and 0.11%, respectively.) This is because the bond issuer is initially considered to be creditworthy and the more time that elapses, the greater the possibility that its financial health will decline. For bonds with a poor credit rating, the probability of default can be a decreasing function of time. (For example, the probabilities that a bond initially in the lowest (CCC/C) category will default during years one, two, three, and four are 28.3%, 10.03%, 5.09%, and 2.94%, respectively.) The reason here is that, for a bond with a poor credit rating, the next year or two may be critical. If the issuer survives this period, its financial health is likely to have improved.

17.2.1 Hazard Rates

From Table 17.1, we can calculate the probability of a CCC/C bond defaulting during the third year as $43.42 - 38.33 = 5.09\%$. We will refer to this as the *unconditional default probability*. It is the probability of default during the third year as seen at time zero. The probability that the bond will survive until the end of year two is $100 - 38.33 = 61.67\%$. The probability that it will default during the third year conditional on no earlier default is therefore $0.0509/0.6167$ or 8.25%.

The 8.25% we have just calculated is a conditional default probability for a one-year time period. When we consider a conditional default probability for a short time period of length Δt , we get a measure known as the *hazard rate* or *default intensity*. The hazard rate, $\lambda(t)$, at time t is defined so that $\lambda(t)\Delta t$ is the probability of default between time t and $t + \Delta t$ conditional on no default between time zero and time t . If $V(t)$ is the cumulative probability of the company surviving to time t (i.e., no default by time t), the unconditional default probability between times t and $t + \Delta t$ is $[V(t) - V(t + \Delta t)]$. The probability of default between times t and $t + \Delta t$ conditional on no earlier default is $[V(t) - V(t + \Delta t)]/V(t)$. Hence

$$\frac{V(t) - V(t + \Delta t)}{V(t)} = \lambda(t)\Delta t$$

or

$$\frac{V(t + \Delta t) - V(t)}{\Delta t} = -\lambda(t)V(t)$$

Taking limits

$$\frac{dV(t)}{dt} = -\lambda(t)V(t)$$

from which

$$V(t) = e^{-\int_0^t \lambda(\tau) d\tau}$$

Defining $Q(t)$ as the probability of default by time t , so that $Q(t) = 1 - V(t)$ gives

$$Q(t) = 1 - e^{-\int_0^t \lambda(\tau) d\tau}$$

or

$$Q(t) = 1 - e^{-\bar{\lambda}(t)t} \quad (17.2)$$

where $\bar{\lambda}(t)$ is the average hazard rate between time zero and time t .

Example 17.2

Suppose that the hazard rate is a constant 1.5% per year. The probability of a default by the end of the first year is $1 - e^{-0.015 \times 1} = 0.0149$. The probability of a default by the end of the second year is $1 - e^{-0.015 \times 2} = 0.0296$. The probabilities of a default by the end of the third, fourth, and fifth years are similarly 0.0440, 0.0582, and 0.0723. The unconditional probability of a default during the fourth year is $0.0582 - 0.0440 = 0.0142$. The probability of default in the fourth year, conditional on no earlier default, is $0.0142 / (1 - 0.0440) = 0.0149$.

17.3 Recovery Rates

When a company goes bankrupt, those who are owed money by the company file claims against the company.⁴ Sometimes there is a reorganization in which these creditors agree to a partial payment of their claims. In other cases, the assets are sold by the liquidator and the proceeds are used to meet the claims as far as possible. Some claims typically have priorities over other claims and are met more fully.

The *recovery rate* for a bond is normally defined as the price at which it trades about one month after default as a percent of its principal amount. Table 17.2 provides historical data on average default rates for different categories of bonds in the United States. These vary from 54.6% from first lien bonds to 22.3% for those ranking after other lenders. (A first lien bond has first claim on collateral; a second lien bond has second claim on collateral. Amongst bonds which are not protected by collateral, senior unsecured bonds rank above senior subordinated bonds, which rank above subordinated bonds, which in turn rank above junior subordinated bonds.)

⁴In the United States, the claim made by a bondholder is the bond's face value plus accrued interest.

Table 17.2 Recovery Rates on Corporate Bonds as a Percent of Face Value, 1983 to 2021, Issuer Weighted

Class	Average Recovery Rate (%)
1st lien bond	54.6
2nd lien bond	44.3
Senior unsecured bond	38.0
Senior subordinated bond	31.0
Subordinated bond	32.1
Junior subordinated bond	22.3

SOURCE: Moody's.

17.3.1 The Dependence of Recovery Rates on Default Rates

In Chapter 7, we saw that one of the lessons from the Global Financial Crisis of 2007–2008 is that the average recovery rate on mortgages is negatively related to the mortgage default rate. As the mortgage default rate increases, foreclosures lead to more houses being offered for sale and a decline in house prices. This in turn results in a decline in recovery rates.

The average recovery rate on corporate bonds exhibits a similar negative dependence on default rates.⁵ In a year when the number of bonds defaulting is low, economic conditions are usually good and the average recovery rate on those bonds that do default might be as high as 60%; in a year when the default rate on corporate bonds is high, economic conditions are usually poor and the average recovery rate on the defaulting bonds might be as low as 30%. The result of the negative dependence is that a bad year for defaults is doubly bad for a lender because it is usually accompanied by a low recovery rate.

17.4 Credit Default Swaps

A credit default swap (CDS) is a derivative that allows market participants to trade credit risks. As indicated in Business Snapshot 17.1, the market for this product grew rapidly until 2007 and then declined. The simplest type of CDS is an instrument that provides insurance against the risk of a default by a particular company. The company is known as the *reference entity* and a default by the company is known as a *credit event*. The buyer of the insurance obtains the right to sell bonds issued by the company for their face value when a credit event occurs and the seller of the insurance agrees to buy the bonds for their face value when a credit event occurs.⁶ The total face value of the bonds that can be sold is known as the credit default swap's *notional principal*.

⁵See E. I. Altman, B. Brady, A. Resti, and A. Sironi, “The Link between Default and Recovery Rates: Theory, Empirical Evidence, and Implications,” *Journal of Business* (November 2005): 2203–2228. The correlation is also discussed in publications by Moody's Investors Service. It finds that the average recovery rate in a year and the non-investment-grade default rate are highly correlated.

⁶The face value (or par value) of a bond is the principal amount that the issuer will repay at maturity if it does not default.

BUSINESS SNAPSHOT 17.1

The CDS Market

In 1998 and 1999, the International Swaps and Derivatives Association developed a standard contract for trading credit default swaps in the over-the-counter market. After that, the market grew very fast. The Bank for International Settlements (BIS) started producing statistics for the size of the credit derivatives market in December 2004. At that time, the total notional principal underlying outstanding contracts was estimated to be about \$6 trillion. It peaked at \$58 trillion in December 2007. But it had fallen to under \$10 trillion by December 2021. Banks and other financial institutions are both buyers and sellers of protection. Banks tend to be net buyers of protection, and insurance companies tend to be net sellers of protection. Banks sometimes use credit default swaps to hedge an exposure to a borrower. It is then the case that the financial institution bearing the credit risk of a loan is different from the financial institution that did the original credit checks.

During the credit turmoil that started in August 2007, regulators became concerned that CDSs were a source of systemic risk. No doubt their concerns arose in part because of the losses experienced by the insurance company AIG. This was a big seller of protection on the AAA-rated tranches created from mortgages (see Chapter 7). The protection proved very costly to AIG and a failure of AIG would have led to big losses elsewhere in the financial system. AIG was bailed out by the United States government in September 2008.

CDSs have come under criticism during the European sovereign debt crisis. Some legislators feel that speculative activity in credit default swap markets has exacerbated the debt problems of countries such as Greece; naked CDS positions on sovereign debt (where credit protection on a country is bought without an underlying exposure) were banned in Europe in 2013.

During 2007 and 2008, trading ceased in many types of credit derivatives, but plain vanilla CDSs, which provide protection against a single company or country defaulting, continued to trade actively (albeit with dramatically increased spreads). This is because their structure is easy to understand compared with, say, ABS CDOs (see Chapter 7).

There was a huge number of CDS contracts outstanding with Lehman Brothers as the reference entity when Lehman Brothers declared bankruptcy in September 2008. The recovery rate (determined by an auction process) was only about eight cents on the dollar, so that the payout to the buyers of protection was equal to about 92% of the notional principal. There were predictions that some sellers of protection would be unable to pay and that further bankruptcies would occur, but on the settlement day (October 21, 2008) everything went smoothly.

The buyer of a CDS makes periodic payments to the seller until the end of the life of the CDS or until a credit event occurs. These payments are usually made in arrears every quarter.

An example will help to illustrate how a typical deal is structured. Suppose that two parties enter into a five-year credit default swap on December 20, 2024. Assume that the notional principal is \$100 million and the buyer agrees to pay 90 basis points per year (quarterly in arrears) for protection against default by the reference entity.

The CDS is shown in Figure 17.1. If the reference entity does not default (that is, there is no credit event), the buyer receives no payoff and pays approximately \$225,000 ($= 0.25 \times 0.0090 \times 100,000,000$) on March 20, June 20, September 20, and December 20 of each of the years 2025, 2026, 2027, 2028, and 2029.⁷ If there is a credit event, a substantial payoff is likely. Suppose that the buyer notifies the seller of a credit event on May 20, 2027 (five months into the third year). If the contract specifies physical settlement, the buyer of protection has the right to sell to the seller of protection bonds issued by the reference entity with a face value of \$100 million for \$100 million. If, as is now usual, there is a cash settlement, a two-stage auction process is used to determine the mid-market value of the cheapest deliverable bond several days after the credit event. Suppose the auction indicates that the cheapest deliverable bond is worth \$35 per \$100 of face value. The cash payoff would be \$65 million.

The regular payments from the buyer of protection to the seller of protection cease when there is a credit event. However, because these payments are made in arrears, a final accrual payment by the buyer is usually required. In our example, where there is a default on May 20, 2027, the buyer would be required to pay to the seller the amount of the annual payment accrued between March 20, 2027, and May 20, 2027 (approximately \$150,000), but no further payments would be required.

The total amount paid per year, as a percent of the notional principal, to buy protection is known as the *CDS spread*. (In our example, the CDS spread is 90 basis points.) Several large banks are market makers in the credit default swap market. For a five-year credit default swap on a company, a market maker might quote: bid 250 basis points, ask 260 basis points. This means that the market maker is prepared to buy protection by paying 250 basis points per year (i.e., 2.5% of the principal per year) and to sell protection for 260 basis points per year (i.e., 2.6% of the principal per year).

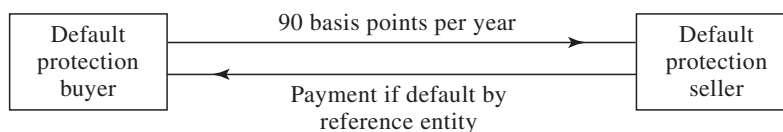


Figure 17.1 Credit Default Swap

⁷The payments are not exactly \$225,000 because of the impact of day count conventions.

BUSINESS SNAPSHOT 17.2

Is a CDS a Flawed Contract?

There is one important difference between credit default swaps and most other over-the-counter derivatives. The other over-the-counter derivatives depend on interest rates, exchange rates, equity indices, commodity prices, and so on. There is no reason to assume that any one market participant has better information than other market participants about these variables.

Credit default swap spreads depend on the probability that a particular company will default during a particular period of time. Arguably, some market participants have more information to estimate this probability than others. A financial institution that works closely with a particular company by providing advice, making loans, and handling new issues of securities is likely to have more information about the creditworthiness of the company than another financial institution that has no dealings with the company. Economists refer to this as an *asymmetric information* problem.

A blatant abuse of the CDS market was reported in the *Financial Times* on January 11, 2018. A U.S. housebuilder, Hovnanian, obtained attractive funding from the hedge fund GSO Capital Partners with the agreement that it would skip an interest payment, thereby manufacturing a technical default. GSO had bought protection on Hovnanian and obtained a big payoff. It is necessary for CDS contracts to find a way of outlawing manufactured credit events of this type if the CDS market is to survive.

Many different companies and countries are reference entities for the CDS contracts traded. Contracts with maturities of five years are most popular, but other maturities such as 1, 2, 3, 7, and 10 years are also traded. Usually, contracts mature on the following standard dates: March 20, June 20, September 20, and December 20. The effect of this is that the actual time to maturity of a contract when it is initiated is close to, but not necessarily the same as, the number of years to maturity that is specified. Suppose you call a dealer on November 15, 2024, to buy five-year protection on a reference entity. The contract would probably last until December 20, 2029. Your first payment would be due on December 20, 2024, and would equal an amount covering the November 15, 2024, to December 20, 2024, period.⁸ After that, payments would be made quarterly.

⁸If the time to the first date is less than one month, the first payment is typically on the second payment date, otherwise it is on the first payment date.

A key aspect of a CDS contract is the definition of a credit event (i.e., a default). Usually a credit event is defined as a failure to make a payment as it becomes due, a restructuring of debt, or a bankruptcy. Restructuring is sometimes excluded in North American contracts, particularly in situations where the yield on the company's debt is high. Potential flaws in the CDS market are discussed in Business Snapshot 17.2. They may account for the market's declining popularity.

17.4.1 The Cheapest-to-Deliver Bond

Usually a CDS specifies that a number of different bonds can be delivered in the event of a default. The bonds typically have the same seniority, but they may not sell for the same percentage of face value immediately after a default.⁹ This gives the holder of a CDS what is known as a *cheapest-to-deliver bond* option. As already mentioned, an auction process is usually used to determine the value of the cheapest-to-deliver bond and, therefore, the payoff to the buyer of protection.

The determination of CDS spreads and the valuation of CDS transactions is discussed in Appendix K.

17.4.2 Credit Indices

Participants in credit markets have developed indices to track credit default swap spreads. In 2004, there were agreements between different producers of indices that led to some consolidation. Two important standard portfolios used by index providers are:

1. CDX NA IG, a portfolio of 125 investment-grade companies in North America
2. iTraxx Europe, a portfolio of 125 investment-grade companies in Europe

These portfolios are updated in March and September each year. Companies that are no longer investment grade are dropped from the portfolios and new investment-grade companies are added.¹⁰

Suppose that the five-year CDX NA IG index is quoted by a market maker as bid 165 basis points, ask 166 basis points. The quotes mean that a trader can buy CDS protection on all 125 companies in the index for 166 basis points per company. Suppose

⁹There are a number of reasons for this. The claim that is made in the event of a default is typically equal to the bond's face value plus accrued interest. Bonds with high accrued interest at the time of default therefore tend to have higher prices immediately after default. Also, the market may judge that in the event of a reorganization of the company some bondholders will fare better than others.

¹⁰In March 2022, the Series 37 iTraxx Europe portfolio and the Series 38 CDX NA IG portfolio were defined. The series numbers indicate that by the end of March 2022 the iTraxx Europe portfolio had been updated 36 times and the CDX NA IG portfolio had been updated 37 times.

an investor wants \$800,000 of protection on each company. The total cost is $0.0166 \times 800,000 \times 125$ or \$1,660,000 per year. (The investor could similarly sell \$800,000 of protection on each of the 125 companies for a total of \$1,650,000 per annum.) When a company defaults, the investor receives the usual CDS payoff and the annual payment is reduced by $1,660,000/125 = \$13,280$. There is an active market in buying and selling CDS index protection for maturities of 3, 5, 7, and 10 years. The maturities for these types of contracts on the index are usually December 20 and June 20. (This means that a “five-year” contract lasts between $4\frac{3}{4}$ and $5\frac{1}{4}$ years.) Roughly speaking, the index is the average of the CDS spreads on the companies in the underlying portfolio.¹¹

17.4.3 The Use of Fixed Coupons

To facilitate trading, the precise way in which CDS and CDS index transactions work is a little more complicated than has been described up to now. In practice, CDS and CDS indices trade like bonds. For each of the standard transactions that trade, a fixed coupon and a recovery rate are specified. The periodic protection payments are equal to the coupon. The CDS spread is quoted in the market. When there is a trade, the present value of the difference between payments equal to the spread and payments equal to the coupon for the remaining life of the transaction is estimated using a specified procedure. If the quoted spread is less than the coupon, the seller of protection pays this present value to the buyer of protection at the time of the trade. If the quoted spread is greater than the coupon, the buyer of protection pays this present value to the seller of protection at the time of the trade. After that, the buyer of protection then pays the coupon times the principal applicable to those companies in the index that have not yet defaulted to the seller.

17.5 Credit Spreads

The credit spread is the extra rate of interest per annum required by investors for bearing a particular credit risk. CDS spreads, which were explained in the previous section, provide one measure of the credit spread. Another is the bond yield spread. This is the amount by which the yield on a corporate bond exceeds the yield on a similar risk-free bond. We now show that the two should be approximately equal.

¹¹More precisely, the index is slightly lower than the average of the credit default swap spreads for the companies in the portfolio. To understand the reason for this, consider a portfolio consisting of two companies, one with a spread of 1,000 basis points and the other with a spread of 10 basis points. To buy protection on the companies would cost slightly less than 505 basis points per company. This is because the 1,000 basis points is not expected to be paid for as long as the 10 basis points and should therefore carry less weight. Another complication for CDX NA IG, but not iTraxx Europe, is that the definition of default applicable to the index includes restructuring whereas the definition for CDS contracts on the underlying companies may not.

17.5.1 CDS Spreads and Bond Yields

A CDS can be used to hedge a position in a corporate bond. Suppose that an investor buys a five-year corporate bond yielding 7% per year for its face value and at the same time enters into a five-year CDS to buy protection against the issuer of the bond defaulting. Suppose that the CDS spread is 200 basis points or 2% per annum. The effect of the CDS is to convert the corporate bond to a risk-free bond (at least approximately). If the bond issuer does not default, the investor earns 5% per year (when the CDS spread is netted against the corporate bond yield). If the bond issuer does default, the investor earns 5% up to the time of the default. Under the terms of the CDS, the investor is then able to exchange the bond for its face value. This face value can be invested at the risk-free rate for the remainder of the five years.

This argument shows that the n -year CDS spread should be approximately equal to the excess of the par yield on an n -year corporate bond over the par yield on an n -year risk-free bond.¹² If it is markedly less than this, an investor can earn more than the risk-free rate by buying the corporate bond and buying protection. If it is markedly greater than this, an investor can borrow at less than the risk-free rate by shorting the corporate bond and selling CDS protection. As will be discussed later, these are not perfect arbitrages, and the extent to which they can be used is in practice influenced by liquidity constraints and other considerations.

17.5.2 The Risk-Free Rate

CDSs provide a direct estimate of the credit spread. To calculate a credit spread from a bond yield, it is necessary to make an assumption about the risk-free rate. When bond yield spreads are quoted by bond traders, the risk-free rate that is used is usually the yield on a Treasury bond of similar maturity. For example, a bond trader might quote the yield on a particular corporate bond as being a spread of 250 basis points over Treasuries.

A number of researchers have compared bond yields to CDS spreads to imply a risk-free rate. This involves matching the maturities of CDSs and bonds and implying a risk-free rate from the arbitrage arguments given above. For example, if the five-year bond yield is 4.7% and the five-year CDS spread is 80 basis points, the implied five-year risk-free rate is 3.9%. The implied risk-free rate has been found to be higher than the Treasury rate. One estimate puts it close to the short-term borrowing rate for a AA-rated company.¹³

¹²The par yield on an n -year bond is the coupon rate per year that causes the bond to sell for its par value (i.e., its face value).

¹³See J. Hull, M. Predescu, and A. White, "The Relationship between Credit Default Swap Spreads, Bond Yields, and Credit Rating Announcements," *Journal of Banking and Finance* 28 (November 2004): 2789–2811.

17.5.3 Asset Swaps

Asset swaps provide a convenient reference point for traders in credit markets because they give direct estimates of the excess of bond yields over a floating reference rate.

To explain how asset swaps work, consider the situation where an asset swap spread for a particular bond is quoted as 150 basis points. There are three possible situations:

1. The bond sells for its par value of 100. The swap then involves one side (Company A) paying the coupon on the bond and the other side (Company B) paying the floating reference rate plus 150 basis points.¹⁴
2. The bond sells below its par value, say, for 95. The swap is then structured so that Company A pays \$5 per \$100 of notional principal at the outset. After that, Company A pays the bond's coupons and Company B pays the floating reference rate plus 150 basis points.
3. The underlying bond sells above par, say, for 108. The swap is then structured so that Company B makes a payment of \$8 per \$100 of principal at the outset. After that, Company A pays the bond's coupons and Company B pays the floating reference rate plus 150 basis points.

The effect of structuring asset swaps in this way is that the present value of the asset swap spread (150 basis points in our example) is the amount by which the price of the corporate bond is exceeded by the price of a similar risk-free bond where the risk-free rate is assumed to be given by the floating reference rate (see Problem 17.16).

17.5.4 CDS–Bond Basis

The CDS–bond basis is the excess of the CDS spread over the bond yield spread for a company.

$$\text{CDS–Bond Basis} = \text{CDS Spread} - \text{Bond Yield Spread}$$

The arbitrage argument given in Section 17.5.1, relating CDS spreads and bond yields, suggests that the CDS–bond basis should be close to zero. In fact, there are a number of reasons why it deviates from zero. For example:

1. The bond may sell for a price that is significantly different from par. (Bond prices above par tend to give rise to a negative basis; bond prices below par tend to give rise to a positive basis.)
2. There is counterparty default risk in a CDS. (This pushes the basis in a negative direction.)
3. There is a cheapest-to-deliver bond option in a CDS. (This pushes the basis in a positive direction.)

¹⁴Note that it is the promised coupons that are exchanged. The exchanges take place regardless of whether the bond defaults.

4. The payoff in a CDS does not include accrued interest on the bond that is delivered. (This pushes the basis in a negative direction.)
5. The restructuring clause in a CDS contract may lead to a payoff when there is no default. (This pushes the basis in a positive direction.)
6. The bond-yield spread might be calculated relative to a “risk-free” reference rate that is not the one used by the market.

Prior to the market turmoil starting in 2007, the basis tended to be positive. For example, De Wit estimates that the average CDS bond basis in 2004 and 2005 was 16 basis points.¹⁵ During the Global Financial Crisis, the basis was at times very negative but, as explained by Bai and Collin-Dufresne (2011), it was difficult for financial institutions to arbitrage between bonds and CDSs because of a shortage of liquidity and other considerations.¹⁶ Since the Global Financial Crisis, the CDS-bond basis has usually been small and negative.

17.6 Estimating Default Probabilities from Credit Spreads

We now discuss how default probabilities can be estimated from credit spreads.

17.6.1 Approximate Calculation

Suppose that a five-year credit spread (CDS spread, bond yield spread, or asset swap spread) for a company is 240 basis points and that the expected recovery rate in the event of a default is 40%. The holder of a corporate bond issued by the company must be expecting to lose 240 basis points (or 2.4% per year) from defaults. Roughly speaking, the credit spread can be considered to be an average loss rate. Given the recovery rate of 40%, this leads to an estimate of the average probability of a default per year over the five-year period, conditional on no earlier default, of $0.024/(1 - 0.4)$ or 4%. In general:

$$\bar{\lambda} = \frac{s(T)}{1 - R} \quad (17.3)$$

where $s(T)$ is the credit spread (which should in theory be expressed with continuous compounding) for a maturity of T , R is the recovery rate, and $\bar{\lambda}$ is the average hazard rate between time zero and time T .

If credit spreads are known for a number of different maturities, the term structure of the hazard rate can be bootstrapped (at least approximately), as the following example illustrates.

¹⁵See J. De Wit, “Exploring the CDS–Bond Basis,” Working Paper no. 104, National Bank of Belgium, 2006.

¹⁶See J. Bai and P. Collin-Dufresne, “The CDS–Bond Basis,” Working Paper, SSRN 2024531, 2011.

Example 17.3

Suppose that the CDS spreads for 3-, 5-, and 10-year instruments are 50, 60, and 100 basis points, respectively, and the expected recovery rate is 60%. The average hazard rate over three years is approximately $0.005/(1 - 0.6) = 0.0125$. The average hazard rate over five years is approximately $0.006/(1 - 0.6) = 0.015$. The average hazard rate over 10 years is approximately $0.01/(1 - 0.6) = 0.025$. From this we can estimate that the average hazard rate between year 3 and year 5 is $(5 \times 0.015 - 3 \times 0.0125)/2 = 0.01875$. The average hazard rate between year 5 and year 10 is $(10 \times 0.025 - 5 \times 0.015)/5 = 0.035$.

17.6.2 A More Exact Calculation

The calculation we have just given works well for CDS spreads. It also works well for bond yield spreads and asset swap spreads when the underlying bond is selling for close to its par value. We now consider a more exact calculation for situations when the underlying bond's price is not close to par.

Suppose that a five-year corporate bond with a principal of 100 provides a coupon of 6% per annum (paid semiannually) and that the yield on the bond is 7% per annum (with continuous compounding). The yield on a similar risk-free bond is 5% (again with continuous compounding). The yields imply that the price of the corporate bond is 95.34 and the price of the risk-free bond is 104.09. The expected loss from default over the five-year life of the bond is therefore $104.09 - 95.34$, or \$8.75. For simplicity, we suppose that the unconditional probability of default per year is the same each year and equal to Q . Furthermore, we assume defaults can happen only at times 0.5, 1.5, 2.5, 3.5, and 4.5 years (immediately before coupon payment dates). Risk-free rates are assumed to be 5% (with continuous compounding) for all maturities and the recovery rate (i.e., recovery as a percent of principal) is assumed to be 40%. (The analysis can be extended so that defaults happen more frequently.)

Table 17.3 calculates the expected loss from defaults in terms of Q . To illustrate the calculations, consider the 3.5-year row in Table 17.3. The expected value of the

Table 17.3 Calculation of Loss from Default on a Bond in Terms of the Default Probabilities per Year, Q

Time (yrs)	Def. Prob.	Recovery Amount (\$)	Default-Free Value (\$)	Loss (\$)	Discount Factor	PV of Expected Loss (\$)
0.5	Q	40	106.73	66.73	0.9753	65.08 Q
1.5	Q	40	105.97	65.97	0.9277	61.20 Q
2.5	Q	40	105.17	65.17	0.8825	57.52 Q
3.5	Q	40	104.34	64.34	0.8395	54.01 Q
4.5	Q	40	103.46	63.46	0.7985	50.67 Q
Total						288.48 Q

Notional Principal = \$100.

default-free bond at time 3.5 years (calculated using the forward risk-free interest rates) is

$$3 + 3e^{-0.05 \times 0.5} + 3e^{-0.05 \times 1.0} + 103e^{-0.05 \times 1.5} = 104.34$$

Given the definition of recovery rates in Section 17.3, the amount recovered if there is a default is 40 so that the loss given default is $104.34 - 40$ or \$64.34. The present value of this loss is 54.01 and the expected loss is therefore 54.01Q.

Table 17.3 shows that the total expected loss is 288.48Q. Setting this equal to the 8.75 expected loss calculated earlier, we obtain a value for Q of $8.75/288.48$, or 3.03%. The calculations we have given assume that the default probability is the same in each year and that defaults take place at just one time during the year. We can extend the calculations to assume that defaults take place more frequently. Also, instead of assuming a constant unconditional probability of default, we can assume a constant hazard rate or assume a particular pattern for the variation of the default probability with time. With several bonds, we can estimate several parameters describing the term structure of default probabilities. Suppose, for example, that we have bonds maturing in 3, 5, 7, and 10 years and we assume a step function for the default probability. We could use the first bond to estimate the default probability per year for the first three years, the second bond to estimate the default probability per year for years 4 and 5, the third bond to estimate the default probability per year for years 6 and 7, and the fourth bond to estimate the default probability per year for years 8, 9, and 10 (see Problems 17.15 and 17.24). This approach is analogous to the bootstrap procedure for estimating the term structure of interest rates in Appendix B.

17.7 Comparison of Default Probability Estimates

We now compare the default probability estimates calculated from historical data with those calculated from credit spreads. Table 17.4 shows:

- The seven-year average cumulative probability of default for companies with different credit ratings. This is taken from data published by S&P. It is the seven-year column of Table 17.1.

Table 17.4 Cumulative Default Probabilities Compared with Credit Spreads

Rating	Cumulative 7-Year Default Probabilities (%), 1981–2020	7-Year Credit Spread (bp) 1996–2007
AA	0.51	35.74
AA	0.49	43.67
A	0.76	68.68
BBB	2.27	127.53
BB	8.89	280.28
B	20.99	481.04
CCC/C	50.75	1,103.70

- The average credit spread for bonds with different credit ratings between December 1996 and June 2007. This is taken from data on bond yields published by Merrill Lynch and an assumption that the reference risk-free rate is 10 basis points less than the seven-year swap rate. The bonds had an average maturity of about seven years.

The table considers data on bond yields only up to the start of the 2008 Global Financial Crisis. During the crisis, credit spreads soared. If that unusual period had been included, the results to be presented, which show that hazard rates calculated from credit spreads are higher than those calculated from historical default probabilities, would become much more extreme.

Table 17.5 shows estimates of the average seven-year hazard rates.¹⁷ Suppose that Q is the seven-year cumulative default probability. From equation (17.2)

$$Q = 1 - e^{-7\bar{\lambda}}$$

where $\bar{\lambda}$ is the average hazard rate so that

$$\bar{\lambda}(7) = -\frac{1}{7} \ln(1 - Q)$$

Consider, for example, an A-rated company. The cumulative seven-year default rate is 0.76% or 0.0076 so that the average hazard rate is

$$-\frac{1}{7} \ln(1 - 0.0076) = 0.00109$$

or 0.109%. The hazard rates implied by credit spreads in Table 17.5 are based on equation (17.3). The recovery rate is assumed to be 40%. Consider again A-rated bonds. The average seven-year credit spread in Table 17.4 is 68.68 basis points or 0.006868. The average seven-year hazard rate is therefore

$$0.006868/(1 - 0.4) = 0.01145$$

Table 17.5 Average Seven-Year Hazard Rates

Rating	Rate from Default Probabilities (%)	Rate Implied by Credit Spreads (%)	Difference
AAA	0.073	0.596	0.523
AA	0.070	0.728	0.658
A	0.109	1.145	1.036
BBB	0.328	2.126	1.798
BB	1.330	4.671	3.341
B	3.366	8.017	4.651
CCC/C	10.118	18.395	8.207

% per annum.

¹⁷The analysis here is an update to that in J. Hull, M. Predescu, and A. White, "Bond Prices, Default Probabilities, and Risk Premiums," *Journal of Credit Risk* 1, no. 2 (Spring 2005): 53–60.

Table 17.5 shows that the hazard rates implied by credit spreads before the 2008 Global Financial Crisis are higher than those calculated from a long period of historical data. The difference between the two hazard rates tends to increase as credit quality declines. From this, one can conclude that investors get well compensated for taking credit risk. The compensation is greater than an actuarial estimate of the cost of defaults. Furthermore, this excess compensation, or excess return, increases as the amount of the credit risk assumed increases.¹⁸

The excess return does not remain constant through time. When credit spreads are high (as they were, for example, during the 2008 Global Financial Crisis), excess returns tend to increase.

17.7.1 Reason for the Difference

Why do we see such big differences between hazard rates estimated from default probabilities and those implied by credit spreads? This is the same as asking why corporate bond traders on average get compensated for more than just the expected cost of defaults.

One reason for the results is that corporate bonds are relatively illiquid and the returns on bonds are higher than they would otherwise be to compensate for this. But this is a small part of what is going on. In normal markets, it explains perhaps 25 basis points of the excess return earned by bond traders. Another possible reason for the results is that the subjective default probabilities of bond traders are much higher than those given in Table 17.1. Bond traders may be allowing for depression scenarios much worse than anything seen in the period covered by their data. However, it is difficult to see how this can explain a large part of the excess return that is observed.

By far the most important reason for the results in Table 17.5 is that bonds do not default independently of each other. Evidence for this is that default rates vary markedly from year to year. Table 9.6 shows that default rates for all rated companies ranged from a low of 0.15% in 1981 to 4.19% in 2009. This year-to-year variation in default rates gives rise to systematic risk (i.e., risk that cannot be diversified away). Bond traders earn an excess expected return for bearing this risk. In this respect, bond traders are no different from equity traders. The average extra return earned by equity traders for bearing systematic risk is often assumed to be 5% or 6% per year. The excess return earned by bond traders is much less than this for high-quality bonds. However, as the bond's credit quality decreases, it becomes more like equity and the excess return earned tends to increase.

What causes default correlation and the resultant systematic risk? One explanation is the economy. Good macroeconomic conditions decrease the probability of default for all companies; bad macroeconomic conditions increase the probability of default for all companies. (In Vasicek's model, which was discussed in Chapter 9, the factor F can be

¹⁸Research identifying a similar difference between the two types of default probability estimates includes J. S. Fons, "The Default Premium and Corporate Bond Experience," *Journal of Finance* 42, no. 1 (March 1987): 81–97, and E. I. Altman, "Measuring Corporate Bond Mortality and Performance," *Journal of Finance* 44, no. 4 (September 1989): 909–922.

BUSINESS SNAPSHOT 17.3

Contagion

Credit contagion is the process whereby a problem in one sector of the world economy leads to problems in other, unrelated sectors. When Russia defaulted on its debt in 1998, there was a flight to quality and credit spreads on all bonds increased. During the Global Financial Crisis that started in 2007, there was a similar flight to quality and again credit spreads increased. The accompanying recession led to a record number of companies defaulting in 2009. In 2011, problems experienced by Greece caused investors to be reluctant to buy the debt of other countries such as Spain, Ireland, Portugal, and Italy. As a result, credit spreads on the debt issued by these countries increased sharply.

The reasons for credit contagion have been debated by researchers. It may be the case that investors become more risk averse when they lose money in one sector. It may be the case that problems in one sector lead investors to become more pessimistic about other, unrelated sectors. Whatever the reason, entities in unrelated sectors are liable to find it more difficult to fund their activities and, as a result, may become more likely to default.

regarded as representing the overall health of the economy.) Another explanation is what is known as “contagion.” This is discussed in Business Snapshot 17.3.

In addition to systematic risk, which we have just talked about, there is nonsystematic (or idiosyncratic) risk associated with each bond. If we were talking about stocks, we would argue that investors can diversify the nonsystematic risk by choosing a portfolio of, say, 30 stocks. They should not therefore demand a risk premium for bearing nonsystematic risk. For bonds the arguments are not so clear cut. Bond returns are highly skewed, with limited upside. (For example, on an individual bond there might be a 99.75% chance of a 7% return in a year, and a 0.25% chance of a –60% return in the year, the first outcome corresponding to no default and the second to default.) The nonsystematic component of this risk is difficult to “diversify away.”¹⁹ It requires tens of thousands of different bonds to be held. In practice, many bond portfolios are far from fully diversified. As a result, bond traders may earn an extra return for bearing nonsystematic risk as well as for bearing the systematic risk mentioned in the previous paragraph.

17.7.2 Valuation vs. Scenario Analysis

In Chapter 10, we explained that the estimates made for market variables can depend on the use to which the estimates are to be put. For valuation it is common to make

¹⁹See J. D. Amato and E. M. Remolona, “The Credit Spread Puzzle,” *BIS Quarterly Review* (December 2003): 51–63.

estimates in a risk-neutral world (i.e., a world where individuals do not require an extra return for bearing risk) and to use risk-neutral valuation. For scenario analysis, real-world estimates are appropriate.

Which estimates of default rates should be used for valuation and scenario analysis? The estimates from credit spreads can be used for valuation the same way the risk-neutral estimates of other variables are used. For example, to value a bond we can use the credit risk estimates of default probabilities and discount at the risk-free rate. However, for scenario analysis it is more appropriate to use the “real-world” estimates obtained from historical data.

17.8 Using Equity Prices to Estimate Default Probabilities

When we use a table such as Table 17.1 to estimate a company’s real-world probability of default, we are relying on the company’s credit rating. Unfortunately, credit ratings are revised relatively infrequently. This has led some analysts to argue that equity prices can provide more up-to-date information for estimating default probabilities.

In 1974, Merton proposed a model where a company’s equity is an option on the assets of the company.²⁰ Suppose, for simplicity, that a firm has one zero-coupon bond outstanding and that the bond matures at time T . Define

V_0 : Value of company’s assets today.

V_T : Value of company’s assets at time T .

E_0 : Value of company’s equity today.

E_T : Value of company’s equity at time T .

D : Amount of debt interest and principal due to be repaid at time T .

σ_V : Volatility of assets (assumed constant).

σ_E : Instantaneous volatility of equity.

If $V_T < D$, it is (at least in theory) rational for the company to default on the debt at time T . The value of the equity is then zero. If $V_T > D$, the company should make the debt repayment at time T and the value of the equity at this time is $V_T - D$. Merton’s model, therefore, gives the value of the firm’s equity at time T as

$$E_T = \max(V_T - D, 0)$$

This shows that the equity of a company is a call option on the value of the assets of the company with a strike price equal to the repayment required on the debt. The Black–Scholes–Merton formula (see Appendix E at the end of this book), assuming no dividend payments, gives the value of the equity today as

$$E_0 = V_0 N(d_1) - De^{-rT} N(d_2) \quad (17.4)$$

²⁰See R. Merton, “On the Pricing of Corporate Debt: The Risk Structure of Interest Rates,” *Journal of Finance* 29 (1974): 449–470.

where

$$d_1 = \frac{\ln(V_0/D) + (r + \sigma_V^2/2)T}{\sigma_V \sqrt{T}}$$

$$d_2 = d_1 - \sigma_V \sqrt{T}$$

and N is the cumulative normal distribution function.

Under Merton's model, the company defaults when the option is not exercised. The probability of this can be shown to be $N(-d_2)$. To calculate this, we require V_0 and σ_V . Neither of these is directly observable. However, if the company is publicly traded, we can observe E_0 . This means that equation (17.4) provides one condition that must be satisfied by V_0 and σ_V . We can also estimate σ_E . From a result in stochastic calculus known as Ito's lemma,

$$\sigma_E E_0 = \frac{\partial E}{\partial V} \sigma_V V_0$$

Here $\partial E/\partial V$ is the delta of the equity. From Appendix E it is $N(d_1)$, so that

$$\sigma_E E_0 = N(d_1) \sigma_V V_0 \quad (17.5)$$

This provides another equation that must be satisfied by V_0 and σ_V . Equations (17.4) and (17.5) provide a pair of simultaneous equations that can be solved for V_0 and σ_V .²¹

Example 17.4

The value of a company's equity is \$3 million and the volatility of the equity is 80%. The debt that will have to be paid in one year is \$10 million. The risk-free rate is 5% per annum. In this case, $E_0 = 3$, $\sigma_E = 0.80$, $r = 0.05$, $T = 1$, and $D = 10$. Solving equations (17.4) and (17.5) yields $V_0 = 12.40$ and $\sigma_V = 0.2123$. The parameter, d_2 , is 1.1408, so that the probability of default is $N(-d_2) = 0.127$ or 12.7%. The market value of the debt is $V_0 - E_0$ or 9.40. The present value of the promised payment on the debt is $10e^{-0.05 \times 1} = 9.51$. The expected loss on the debt is therefore $(9.51 - 9.40)/9.51$ or about 1.2% of its no-default value.

The basic Merton's model we have just presented has been extended in a number of ways. For example, one version of the model assumes that a default occurs whenever the value of the assets falls below a barrier level. Another allows payments on debt instruments to be required at more than one time. Hull et al. (2004) discuss how implied volatilities can be used in conjunction with Merton's model.²²

²¹To solve two non-linear equations of the form $F(x, y) = 0$ and $G(x, y) = 0$, we can use the Solver routine in Excel to find the values of x and y that minimize $[F(x, y)]^2 + [G(x, y)]^2$.

²²See J. Hull, I. Nelken, and A. White, "Merton's Model, Credit Risk, and Volatility Skews," *Journal of Credit Risk* 1, no. 1 (2004): 1–27.

17.8.1 *Performance of the Model*

How well do the default probabilities produced by Merton's model and its extensions correspond to actual default experience? The answer is that Merton's model and its extensions produce a good ranking of default probabilities (risk-neutral or real-world). This means that a monotonic transformation can be estimated to convert the probability of default output from Merton's model into a good estimate of either the real-world or risk-neutral default probability. Moody's KMV and Kamakura provide a service that transforms a default probability produced by Merton's model into a real-world default probability. Other organizations have used Merton's model to estimate credit spreads, which as we have explained, are closely linked to risk-neutral default probabilities. The default probability, $N(-d_2)$, is in theory a risk-neutral default probability because it is calculated from an option pricing model. It may seem strange for Moody's KMV and Kamakura to use it to estimate a real-world default probability. Given the nature of the calibration process we have just described, the underlying assumption is that the rankings of risk-neutral default probabilities, real-world default probabilities, and default probabilities produced by Merton's model are all the same.

17.8.2 *Real-World vs. Risk-Neutral Default Probabilities*

Merton's model provides a way of understanding why default probabilities are higher in the risk-neutral world than in the real world. In a risk-neutral world, the expected growth rate of the value of the company's assets is the risk-free rate. In the real world, the growth rate of the company's assets is usually higher than this (reflecting a risk premium demanded by the market). The probability of the value of the assets dropping below the face value of the debt at a future time is therefore higher in the risk-neutral world than in the real world.

17.8.3 *Distance to Default*

The term *distance to default* has been coined to describe the output from Merton's model. This is the number of standard deviations the asset price must change for default to be triggered T years in the future. One definition of distance to default is d_2 or

$$\frac{\ln V_0 - \ln D + (r - \sigma_V^2/2)T}{\sigma_V \sqrt{T}}$$

As the distance to default declines, the company becomes more likely to default. In Example 17.4, the one-year distance to default is 1.14 standard deviations.

Summary

The estimation of default probabilities and recovery rates is an important activity for risk managers. If a company has issued publicly traded debt, credit ratings provide one source

of information. Rating agencies such as S&P provide extensive statistics on default rates for companies that have been awarded particular credit ratings. The recovery rate is the value of a bond shortly after default as a percentage of its face value.

There are a number of sources of information about credit spreads. The credit default swap (CDS) market is one such source. A CDS is an instrument where one company buys from another company protection against a third company or country (the reference entity) defaulting on its obligations. The CDS spread is the amount paid per year for protection as a percentage of the notional principal. Two other sources of a company's credit spreads are the excess of yields on bonds issued by the company over the risk-free rate and asset swap spreads. Default probabilities can be implied from credit spreads and an assumption about recovery rates.

The default probabilities that are based on historical data, such as those produced by rating agencies, are termed real-world or physical default probabilities. Those based on credit spreads are risk-neutral default probabilities. Risk-neutral default probabilities are higher than real-world default probabilities. Risk-neutral default probabilities should be used for valuation. Real-world default probabilities should be used for scenario analysis. Either type of probability can be calculated using a model developed by Robert Merton in 1974, provided that the model is calibrated appropriately.

Further Reading

- Altman, E. I. "Measuring Corporate Bond Mortality and Performance." *Journal of Finance* 44 (1989): 902–922.
- Duffie, D., and K. Singleton. "Modeling Term Structures of Defaultable Bonds." *Review of Financial Studies* 12 (1999): 687–720.
- Fons, J. S. "The Default Premium and Corporate Bond Experience." *Journal of Finance* 42, no. 1 (March 1987): 81–97.
- Hull, J., M. Predescu, and A. White. "Bond Prices, Default Probabilities, and Risk Premiums." *Journal of Credit Risk* 1, no. 2 (Spring 2005): 53–60.
- Hull, J., M. Predescu, and A. White. "Relationship between Credit Default Swap Spreads, Bond Yields, and Credit Rating Announcements." *Journal of Banking and Finance* 28 (November 2004): 2789–2811.
- Kealhofer, S. "Quantifying Credit Risk I: Default Prediction." *Financial Analysts Journal* 59, no. 1 (2003): 30–44.
- Kealhofer, S. "Quantifying Credit Risk II: Debt Valuation." *Financial Analysts Journal* 59, no. 3 (2003): 78–92.
- Litterman, R., and T. Iben. "Corporate Bond Valuation and the Term Structure of Credit Spreads." *Journal of Portfolio Management* (Spring 1991): 52–64.
- Merton, R. C. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance* 29 (1974): 449–470.
- Rodriguez, R. J. "Default Risk, Yield Spreads, and Time to Maturity." *Journal of Financial and Quantitative Analysis* 23 (1988): 111–117.

Practice Questions and Problems **(Answers at End of Book)**

- 17.1 What are the investment grade ratings used by S&P.
- 17.2 What is meant by a hazard rate?
- 17.3 Calculate the average hazard rate for a B-rated company during the first year from the data in Table 17.1.
- 17.4 Calculate the average hazard rate for a BB-rated company during the third year from the data in Table 17.1.
- 17.5 A credit default swap requires a semiannual payment at the rate of 60 basis points per year. The principal is \$300 million and the credit default swap is settled in cash. A default occurs after four years and two months, and the calculation agent estimates that the price of the cheapest deliverable bond is 40% of its face value shortly after the default. List the cash flows and their timing for the seller of the credit default swap.
- 17.6 Explain the two ways a credit default swap can be settled.
- 17.7 Explain the difference between the default probabilities calculated from historical default experience and those calculated from credit spreads.
- 17.8 What is the formula relating the payoff on a CDS to the notional principal and the recovery rate?
- 17.9 The spread between the yield on a three-year corporate bond and the yield on a similar risk-free bond is 50 basis points. The recovery rate is 30%. Estimate the implied average hazard rate per year over the three-year period.
- 17.10 The spread between the yield on a five-year bond issued by a company and the yield on a similar risk-free bond is 80 basis points. Assuming a recovery rate of 40%, estimate the implied average hazard rate per year over the five-year period. If the spread is 70 basis points for a three-year bond, what do your results indicate about the implied average hazard rate in years 4 and 5?
- 17.11 Should researchers use default probabilities calculated from historical data or from credit spreads for (a) calculating credit value at risk and (b) adjusting the price of a derivative for default?
- 17.12 How are recovery rates usually defined?
- 17.13 Verify that the numbers in the second column of Table 17.5 are consistent with the numbers in the second column of Table 17.4.
- 17.14 A four-year corporate bond provides a coupon of 4% per year payable semiannually and has a yield of 5% expressed with continuous compounding. The risk-free yield curve is flat at 3% with continuous compounding. Assume that defaults can take place at the end of each year (immediately before a coupon or principal payment) and the recovery rate is 30%. Estimate the risk-neutral default probability on the assumption that it is the same each year using the approach in Table 17.3.
- 17.15 A company has issued three- and five-year bonds, each of which has a coupon of 4% per annum payable annually. The yields on the bonds (expressed with continuous compounding) are 4.5% and 4.75%, respectively. Risk-free interest rates are

- 3.5% with continuous compounding for all maturities. The recovery rate is 40%. Defaults can take place halfway through each year. The unconditional risk-neutral default rates per year are Q_1 for years 1 to 3 and Q_2 for years 4 and 5. Estimate Q_1 and Q_2 .
- 17.16 Suppose that in an asset swap, B is the market price of the bond per dollar of principal, B^* is the default-free value of the bond per dollar of principal, and V is the present value of the asset swap spread per dollar of principal. Show that $V = B^* - B$.
- 17.17 Show that, under Merton's model in Section 17.8, the credit spread on a T -year zero-coupon bond is $-\ln[N(d_2) + N(-d_1)/L]/T$ where $L = De^{-rT}/V_0$.
- 17.18 The value of a company's equity is \$2 million and the volatility of its equity is 50%. The debt that will have to be repaid in one year is \$5 million. The risk-free interest rate is 4% per annum. Use Merton's model to estimate the probability of default. (*Hint:* The Solver function in Excel can be used for this question.)
- 17.19 A five-year credit default swap entered into on June 20, 2023, requires quarterly payments at the rate of 400 basis points per year. The principal is \$100 million. A default occurs after four years and two months. The auction process finds the price of the cheapest deliverable bond to be 30% of its face value. List the cash flows and their timing for the seller of the credit default swap.
- 17.20 "The position of a buyer of a credit default swap is similar to the position of someone who is long a risk-free bond and short a corporate bond." Explain this statement.
- 17.21 Why is there a potential asymmetric information problem in credit default swaps?
- 17.22 Suppose that the risk-free yield curve is flat at 6% with continuous compounding and a five-year bond with a coupon of 5% (paid semiannually) sells for 90.00. How much would the bond be worth if it were a risk-free bond? What is the present value of the expected loss from defaults? How would an asset swap on the bond be structured? What is the asset swap spread that would be calculated in this situation?

Further Questions

- 17.23 Suppose that a three-year corporate bond provides a coupon of 7% per year payable semiannually and has a yield of 5% (expressed with semiannual compounding). The yield for all maturities on risk-free bonds is 4% per annum (expressed with semiannual compounding). Assume that defaults can take place every six months (immediately before a coupon payment) and the recovery rate is 45%. Estimate the default probabilities assuming (a) the unconditional default probabilities are the same on each possible default date and (b) the default probabilities conditional on no earlier default are the same on each possible default date.
- 17.24 A company has issued one- and two-year bonds providing 8% coupons, payable annually. The yields on the bonds (expressed with continuous compounding)

are 6.0% and 6.6%, respectively. Risk-free rates are 4.5% for all maturities. The recovery rate is 35%. Defaults can take place halfway through each year. Estimate the risk-neutral default rate each year.

- 17.25 The value of a company's equity is \$4 million and the volatility of its equity is 60%. The debt that will have to be repaid in two years is \$15 million. The risk-free interest rate is 6% per annum. Use Merton's model to estimate the expected loss from default, the probability of default, and the recovery rate (as a percentage of the no-default value) in the event of default. (*Hint:* The Solver function in Excel can be used for this question.)

Chapter 18

xVAs

As explained in Chapter 6, there are two ways in which over-the-counter derivatives transactions are cleared: through central clearing parties (CCPs) and through bilateral agreements. As a result of regulations introduced since the Global Financial Crisis, transactions between financial institutions in standardized derivatives have to be cleared through CCPs. Transactions in nonstandard derivatives between financial institutions are cleared bilaterally, but regulations require a great deal of collateral to be posted so that there is very little credit risk. This chapter focuses on bilaterally cleared over-the-counter derivatives transactions. These include transactions with non-financial end users and transactions with other counterparties that were entered into before the new rules came into effect.

Assessing the credit risk for derivatives transactions is much more complicated than assessing the credit risk for loans because the future exposure (i.e., the amount that could be lost in the event of a default) is not known. If a bank makes a \$10 million uncollateralized five-year loan to a client with repayment of principal at the end, the bank knows that its exposure is approximately \$10 million at all times during the five-year period. If the bank instead enters into a five-year uncollateralized interest rate swap with the client, the future exposure is much less certain. This is because the future value of the swap depends on movements in interest rates. If the value of the interest rate swap to the bank becomes positive, the exposure is equal to the value of the swap (because this is what the bank could lose in the event of a counterparty default). If the value becomes

negative, the exposure is zero (because in that case the bank would not lose anything in the case of a counterparty default).

This chapter discusses what are known as the xVAs. These quantify different aspects of the costs (or benefits) of the credit risks associated with derivatives. The most important xVA is credit value adjustment (CVA), which is the present value of the expected loss from default by the derivatives dealer. Debit (or debt) value adjustment is the present value of the expected gain to the dealer (loss to the counterparty) from a default by the dealer. Examples of other xVAs are funding value adjustment (FVA), margin value adjustment (MVA), and capital value adjustment (KVA).

18.1 Credit Exposure on Derivatives

We start by quickly reviewing the nature of a dealer's exposure to a counterparty created by derivatives transactions. Consider first the simple case where there is only one derivatives transaction outstanding between a derivatives dealer and a counterparty. We assume that the transaction is cleared bilaterally and no collateral is posted by either the dealer or the counterparty.

Three possible situations can be distinguished:

1. The derivative is always a liability to the dealer.
2. The derivative is always an asset to the dealer.
3. The derivative can become either an asset or a liability to the dealer.

An example of a derivative in the first category is a short position in an option; an example in the second category is a long position in an option; an example in the third category is a forward contract or an interest rate swap transaction.

In the first case, the dealer has no credit exposure to the counterparty. If the counterparty defaults (e.g., because it is bankrupt), the transaction is terminated.¹ The dealer will be required to pay the negative value of the derivative (adjusted to reflect the bid–ask spread costs that the dealer will incur in replacing the transaction) to the estate of the defaulting counterparty. There should therefore be neither a gain nor a loss to the dealer.

In the second case, the dealer always has credit exposure to the counterparty. If the counterparty defaults, the dealer is liable to experience a loss. The dealer is an unsecured creditor for the value of the derivative (adjusted to reflect the bid–ask spread costs that the dealer will incur in replacing the transaction). The amount of the loss depends on the value of the derivative at the time of the bankruptcy and the recovery made.

The third case is more complicated. The dealer may or may not have credit exposure to the counterparty in the future. If the counterparty defaults when the value of the derivative is positive to the dealer, the dealer is an unsecured creditor, as in the second case. If the counterparty defaults when the value is negative to the dealer, no loss or gain to the dealer results, as in the first case.

¹Note that, because the derivative is an asset to the counterparty, it must be the case that, if the counterparty goes bankrupt, it has run into financial difficulties for reasons unrelated to the derivative.

When calculating their exposures, dealers typically do not allow for the bid–ask spread they would face in replacing the transaction. In the simple example we have been considering, the dealer’s net exposure is therefore

$$\max(V, 0)$$

where the variable, V , is the market value of the derivative at the time of the default.

When there are many derivatives transactions outstanding between the dealer and the counterparty, the bilateral agreement between the dealer and its counterparties usually means that they are netted (see Section 6.2.4). The dealer’s exposure at any given time, assuming no collateral is posted, is therefore $\max(V, 0)$, with V equal to the net mark-to-market value of all outstanding transactions.

In calculating its credit risk, the dealer is interested in both its credit exposure today and what that exposure might be in the future. In the next few sections, we focus on CVA, the quantification of the cost to the dealer.

18.2 CVA

A dealer calculates what is known as a *credit value adjustment* or CVA for each counterparty with which it has bilaterally cleared OTC derivatives.² This is an estimate of its expected loss from a default by the counterparty. CVA reduces the value of the derivatives on the balance sheet, and an increase (decrease) in the total CVA during a period leads to a decrease (or increase) in profits for the period.

Suppose that T is the life of the longest derivative outstanding with a counterparty. To calculate CVA, the period of time between time zero (today) and time T is divided into a number of intervals (e.g., zero to one month, one to three months, etc.).

Suppose that the i th time interval runs from t_{i-1} to t_i ($t_0 = 0$) and there are n intervals. Define

q_i : the probability of a loss from a default by the counterparty during the i th time interval.

As explained in Chapter 17, because we are engaged in valuation this should be the “risk-neutral” probability of default calculated from credit spreads.

ν_i : the present value of the expected net exposure of the dealer to the counterparty (after taking account of any collateral posted by the counterparty) at the midpoint of the i th time interval, conditional on a default

R : the estimated recovery rate of the dealer in the event of a default by the counterparty when it is an unsecured creditor

²For further reading on CVA and the way it is calculated, see J. Hull and A. White, “CVA and Wrong Way Risk,” *Financial Analysts Journal* 65, no. 5 (September/October 2012): 58–69; E. Canabarro and D. Duffie, “Measuring and Marking Counterparty Risk,” Chapter 9 in *Asset/Liability Management for Financial Institutions*, ed. L. Tilman (New York: Institutional Investor Books, 2003); E. Picault, “Calculating and Hedging Exposure, CVA, and Economic Capital,” in *Counterparty Credit Risk Modeling*, ed. M. Pykhtin (London: Risk Books, 2005); and J. Gregory, *Counterparty Credit Risk: The New Challenge for Financial Markets*, 2nd ed. (Hoboken, NJ: John Wiley & Sons, 2012).

Assuming that the net exposure is independent of the probability of default, the present value of the expected loss from a default during the i th interval can be estimated as

$$(1 - R)q_i v_i$$

and the total expected loss is

$$\text{CVA} = \sum_{i=1}^n (1 - R)q_i v_i \quad (18.1)$$

The q_i are estimated from the counterparty's credit spreads, as discussed in Chapter 17. Suppose that s_i is an estimate of the counterparty's credit spread for a maturity of t_i . From equation (17.3), an estimate of the average hazard rate between time zero and time t_i is

$$\bar{\lambda}_i = \frac{s_i}{1 - R} \quad (18.2)$$

and the probability of no default between times zero and t_i is

$$e^{-\bar{\lambda}_i t_i}$$

so that

$$q_i = e^{-\bar{\lambda}_{i-1} t_{i-1}} - e^{-\bar{\lambda}_i t_i}$$

or, substituting from equation (18.2)

$$q_i = \exp\left(-\frac{s_{i-1} t_{i-1}}{1 - R}\right) - \exp\left(-\frac{s_i t_i}{1 - R}\right) \quad (18.3)$$

The v_i are usually calculated using Monte Carlo simulation. The market variables determining the future value of the transactions that the dealer has with the counterparty are simulated in a risk-neutral world between time zero and time T . On each simulation trial the exposure of the dealer to the counterparty at the midpoint of each interval is calculated. These exposures are averaged and v_i is set equal to the present value of the average exposure at the midpoint of the i th interval. Dealers may have transactions with thousands of counterparties so that the calculation of the v_i for all of them can be computationally quite time consuming.

18.2.1 Collateral and Cure Periods

Collateral agreements must be incorporated into the calculation of the v_i . A collateral agreement between two sides is generally in the credit support annex (CSA) of their ISDA master agreement. This gives the details on how the collateral posted by each side is calculated and specifies haircuts on noncash collateral. Suppose that C is defined as the collateral posted by the counterparty at the time of a default. (If C is negative, $-C$ is the collateral posted by the dealer with the counterparty at the time of the default.) The exposure, E , of the dealer to the counterparty is in all situations

$$E = \max(V - C, 0) \quad (18.4)$$

where V is the market value of outstanding transactions to the dealer at the time of the default.

Equation (18.4) reflects the fact that collateral posted by the counterparty can be used to reduce (and possibly eliminate) an exposure from a positive V . If the dealer has posted collateral with the counterparty ($C < 0$), to the extent that it is greater than the value $-V$ of the transactions to the counterparty, there is an exposure. (As mentioned, exposure is usually defined in terms of the mid-market value of transactions so that bid-ask spreads are not taken into account.)

It is usually the case that a period of time elapses between the time when a counterparty stops posting collateral and the close-out of transactions. This period of time is known as the *cure period* (or *margin period of risk*). It is typically assumed to be 10 days. The effect of the cure period is that the collateral at the time of a default does not reflect the value of the portfolio at the time of the default. It reflects the value 10 days earlier.

Suppose that the midpoint of the i th interval is t_i^* so that $t_i^* = (t_{i-1} + t_i)/2$ and c is the cure period. The Monte Carlo simulation to calculate the ν_i must be structured so that the value of the derivatives portfolio with the counterparty is calculated at times $t_i^* - c$ as well as at time t_i^* ($i = 1, 2, \dots, n$). On each simulation trial the value at time $t_i^* - c$ is used in conjunction with the CSA to calculate the collateral available at time t_i^* . The net exposure at time t_i^* is then calculated using equation (18.4).³ The calculations are illustrated in Example 18.1.

Example 18.1

Suppose there is a two-way zero threshold collateral agreement between a bank and its counterparty. This means that the total collateral to be posted by one side with the other side on a day is specified as $\max(V, 0)$, where V is the value of the outstanding transactions to the other side. The cure period is assumed to be 10 days. This means that the collateral posted by one side at the time of a default can be assumed to equal $\max(V, 0)$ where V is the value to the other side 10 days earlier. Suppose that time τ is the midpoint of one of the intervals used in the bank's CVA calculation.

1. On a particular simulation trial, the value of outstanding transactions to the bank at time τ is 50 and their value 10 days earlier is 45. The calculation assumes that the bank has collateral worth 45 in the event of a default at time τ . The bank's exposure is therefore 5, the uncollateralized value it has in the derivatives transactions.
2. On a particular simulation trial, the value of outstanding transactions to the bank at time τ is 50 and their value 10 days earlier is 55. In this case, it is assumed that the bank will have adequate collateral and its exposure is zero.
3. On a particular simulation trial, the value of outstanding transactions to the bank at time τ is -50 and the value 10 days earlier is -45 . In this case, the bank is assumed to

³It is assumed that the non-defaulting party does not post any collateral or return any collateral during the cure period.

have posted less than 50 of collateral with the counterparty in the event of a default at time τ and its exposure is zero.

4. On a particular simulation trial, the value of outstanding transactions to the bank at time τ is -50 and the value 10 days earlier is -55 . In this case, it is assumed that 55 of the bank's collateral is held by the counterparty 10 days before time τ and, in the event of a default at time τ , none of it is returned. The bank's exposure is therefore 5, the excess collateral it has posted.

18.2.2 Peak Exposure

In addition to calculating CVA, dealers usually calculate the *peak exposure* at the midpoint of each interval. This is a high percentile of the exposures given by the Monte Carlo simulation trials. Suppose for example that the percentile chosen is 97.5% and there are 10,000 Monte Carlo simulation trials. The peak exposure at time t_i^* is the 250th highest exposure recorded at that time. The *maximum peak exposure* is the maximum of the peak exposures across all the t_i^* .

There is a theoretical issue here (which is often ignored in practice). To calculate CVA, we simulate the behavior of market variables in the risk-neutral world and discount at the risk-free rate. (As explained in Chapter 10, it is correct to do this when a valuation is being carried out.) When we calculate peak exposure, we are carrying out a scenario analysis. We are in effect asking, "How bad can our exposure to the counterparty get in the future?" For this purpose we should in theory simulate the behavior of market variables in the real world, not the risk-neutral world.⁴

18.2.3 Downgrade Triggers

Netting and collateral agreements are important ways in which credit risk is reduced in bilaterally cleared derivatives transactions. Sometimes the credit support annex to an ISDA master agreement also includes a downgrade trigger. This is a clause stating that if the credit rating of one side falls below a certain level, the other side has the right to ask for more collateral or to terminate transactions.

AIG provides an example of the operation of a downgrade trigger. Many of AIG's transactions stated that AIG did not have to post collateral provided its credit rating remained above AA. However, once it was downgraded below AA, collateral was required. AIG was downgraded below AA by all three rating agencies on September 15, 2008. This led to collateral calls, which AIG was unable to meet. Its bankruptcy was avoided by a massive government bailout.

Other companies that have run into problems because of downgrade triggers include Enron, Xerox, and Ambac. Downgrade triggers do not provide protection to a company's

⁴When the peak exposure is being calculated for times that are not far in the future, the difference between simulating in the real world and in the risk-neutral world is usually quite small.

counterparties from a big jump in a company's credit rating (for example, from AA to default). Also, downgrade triggers work well for a company's counterparties only when the company is making relatively little use of them. If a company has many downgrade triggers in its contracts, an AIG-type situation can arise where a downgrade below a specified credit rating can lead to huge cash demands on the company. If these cash demands cannot be met, an immediate bankruptcy follows.

18.3 The Impact of a New Transaction

When a new derivatives transaction is being negotiated by a dealer with a counterparty, its incremental effect on CVA may influence the terms offered. If the value of the new transaction is positively correlated with other transactions entered into by the dealer with the counterparty, the incremental effect on CVA is likely to be positive. If this correlation is negative, the new transaction can have the effect of reducing CVA.

Suppose, for example, that a dealer has one transaction outstanding with a counterparty: a five-year forward foreign currency transaction where the counterparty is buying the currency. If the counterparty then expresses interest in entering into a three-year forward foreign currency transaction, the competitiveness of the dealer's quote might depend on which side of the transaction the counterparty wants to take. If the counterparty wants to buy the currency, the new transaction would have the effect of increasing CVA (making the transaction less attractive to the dealer). On the other hand, if the counterparty wants to sell the currency, it would be netted with the existing transaction and have the effect of reducing CVA (making the transaction more attractive to the dealer). These CVA/netting considerations suggest that when transactions are cleared bilaterally and are not fully collateralized, a company will sometimes get the most favorable quotes from a dealer with whom it already has transactions outstanding rather than from a dealer with whom it has done no previous business.

Calculating CVA can be computationally intensive. Often dealers have hundreds or thousands of transactions outstanding with a counterparty.⁵ Calculating the incremental effect of a new transaction on CVA by recomputing CVA is not usually feasible. Luckily there is a computationally efficient approach for calculating incremental CVA.

When the CVA calculations in Section 18.2 are carried out, the paths followed by market variables on each simulation trial and the value of the portfolio on each simulation trial are stored.⁶ When a potential new transaction is being considered, its value at future times is calculated for the values of the market variables that were obtained on the simulation trials. This gives the incremental impact of the transaction on the future

⁵Lehman Brothers, for example, had about 1.5 million derivatives transactions with about 8,000 counterparties at the time of its failure.

⁶To be precise, the values of all market variables at the t_i^* and $t_i^* - c$ for $1 \leq i \leq n$ and the MTM portfolio values at these times are stored until the next time CVA is calculated (at which time they can be deleted).

portfolio value for each of the last-used Monte Carlo simulation trials at all future times. The incremental impact on the exposure at each time for each Monte Carlo trial can then be calculated. From this, the incremental effect on the average exposure at each time can be calculated and equation (18.1) can then be used to calculate the incremental effect on CVA.

To illustrate this, suppose that a portfolio with a counterparty depends only on the price of gold and that the price of gold at time 2.5 years on the 545th simulation trial when CVA is calculated is \$1,572, with the value of the portfolio to the dealer being \$2.4 million. Assuming no collateral is posted, this is also the value of the exposure. If 2.5 years is the midpoint of the 20th time step, this means that ν_{20} is the present value of \$2.4 million received at time 2.5 years. We suppose that this is \$2.3 million.

Suppose that, shortly after CVA is calculated, a potential new transaction with the counterparty, dependent on the price of gold, is contemplated. This transaction is valued at all times for the paths followed by the price of gold on all simulation trials. Suppose that, on the 545th simulation trial, it is calculated that the value of the new transaction will be $-\$4.2$ million at time 2.5 years (when the price of gold is \$1,572). This means that the portfolio value reduces from \$2.3 million to $-\$1.9$ million at time 2.5 years on the 545th simulation trial as a result of the new transaction. This reduces the exposure to zero so that the new ν_{20} is zero. The new transaction therefore has the effect of reducing ν_{20} by \$2.3 million. Similar calculations are carried out for all simulation trials and all times. Define $\Delta\nu_i$ as the average change in ν_i across all the simulation trials. The incremental effect on CVA of the new transaction is estimated as

$$\sum_{i=1}^n (1 - R)q_i \Delta\nu_i$$

18.4 CVA Risk

A dealer has one CVA for each counterparty. These CVAs can themselves be regarded as derivatives. They are particularly complex derivatives. Indeed, the CVA for a counterparty is more complex than any of the transactions between the dealer and the counterparty because it is contingent on the net value of all the transactions between the dealer and the counterparty.

When CVA increases (decreases), the income reported by a derivatives dealer decreases (increases). For this reason, many dealers consider it prudent to try to hedge CVAs in the same way that they hedge other derivatives. This means that they must calculate the Greek letters (delta, gamma, vega, etc.) discussed in Chapter 15.

The variables affecting the ν_i are market variables such as interest rates, exchange rates, commodity prices, and so on. Calculating Greek letters for these is liable to be computationally quite time consuming. For example, to calculate the delta of CVA with respect to an exchange rate, it is necessary to make a small change to the exchange rate

and recompute CVA. A technique known as adjoint differentiation can be used to ease the computational burden.⁷

The variables affecting the q_i are the credit spreads of the counterparty for different maturities. From equation (18.3)

$$q_i = \exp\left(-\frac{s_{i-1}t_{i-1}}{1-R}\right) - \exp\left(-\frac{s_it_i}{1-R}\right)$$

From equation (18.1)

$$\text{CVA} = \sum_{i=1}^n (1-R)q_i\nu_i$$

Using a delta/gamma approximation, the change in CVA resulting from a small parallel shift, Δs , in the term structure of credit spreads (with all the market variables determining the ν_i being assumed to remain fixed) is therefore

$$\begin{aligned} \Delta(\text{CVA}) &= \sum_{i=1}^n \left[t_i \exp\left(-\frac{s_it_i}{1-R}\right) - t_{i-1} \exp\left(-\frac{s_{i-1}t_{i-1}}{1-R}\right) \right] \nu_i \Delta s \\ &\quad + \frac{1}{2(1-R)} \sum_{i=1}^n \left[t_{i-1}^2 \exp\left(-\frac{s_{i-1}t_{i-1}}{1-R}\right) - t_i^2 \exp\left(-\frac{s_it_i}{1-R}\right) \right] \nu_i (\Delta s)^2 \end{aligned} \quad (18.5)$$

This can be calculated without difficulty once the ν_i are known.

Regulators require dealers to use this equation to incorporate the risks arising from changes in credit spreads into market risk capital calculations.⁸

18.5 Wrong-Way Risk

Up to now we have assumed that the probability of default is independent of the exposure. A situation where there is a positive dependence between the two, so that the probability of default by the counterparty tends to be high (low) when the dealer's exposure to the counterparty is high (low), is referred to as wrong-way risk. A situation where there is negative dependence, so that the probability of default by the counterparty tends to be high (low) when the dealer's exposure to the counterparty is low (high), is referred to as right-way risk.

A subjective evaluation of the amount of wrong-way or right-way risk in transactions with a counterparty requires a good knowledge of the counterparty's business, in particular the nature of the risks facing the business. It also requires knowledge of the

⁷See, for example, M. Giles and P. Glasserman, "Smoking Adjoints: Fast Monte Carlo Greeks," *Risk* 19, no. 1 (2006): 88–92; and M. Henrard, "Adjoint Algorithmic Differentiation: Calibration and the Implicit Function Theorem," *Journal of Computational Finance* 17, no. 4 (2014): 37–47.

⁸See Basel Committee on Banking Supervision, "Review of Credit Valuation Adjustment Risk Framework," October 1, 2015.

transactions the counterparty has entered into with other dealers. The latter is difficult to know with any precision.

One situation in which wrong-way risk tends to occur is when a counterparty is using a credit default swap to sell protection to the dealer. When a dealer buys protection from the counterparty and the credit spread of the reference entity increases, the value of the protection to the dealer becomes positive. However, because the credit spreads of different companies tend to be correlated, it is likely that the credit spread of the counterparty has increased so that the counterparty's calculated probability of default has also increased. Similarly, right-way risk tends to occur when a counterparty is buying credit protection from the dealer.

A situation in which a company is speculating by entering into many similar trades with one or more dealers is likely to lead to wrong-way risk for these dealers. This is because the company's financial position and therefore its probability of default is likely to be affected adversely if the trades move against the company.

If a company enters into transactions with a dealer to partially hedge an existing exposure, there should in theory be right-way risk. This is because, when the transactions move against the counterparty, it will be benefiting from the unhedged portion of its exposure so that its probability of default should be relatively low.⁹

A simple way of dealing with wrong-way risk is to use what is termed the "alpha" multiplier to increase ν_i in the model in equation (18.1) where the ν_i and q_i are assumed to be independent. The effect of this is to increase CVA by the alpha multiplier. Regulators set alpha equal to 1.4, but allow banks to use their own models, with a floor for alpha of 1.2. This means that, at minimum, the CVA has to be 20% higher than that given by the model where the ν_i and q_i are assumed to be independent. If a bank does not have its own model for wrong-way risk, it has to be 40% higher. Estimates of alpha reported by banks range from 1.07 to 1.10.

Some models have been developed to capture the dependence of the probability of default on the exposure. For example, Hull and White (2012) propose a simple model where the hazard rate at time t is a function of variables observable at that time.¹⁰ Their parameter describing the extent of the dependence can be either estimated subjectively or estimated by relating past credit spreads for the counterparty to what the value of the current portfolio would have been in the past. Implementing the model involves relatively minor modifications to the calculations outlined in Section 18.2.

18.6 DVA

Debit value adjustment (DVA) is the mirror image of CVA.¹¹ Whereas CVA is the expected cost to the dealer from a possible default by the counterparty, DVA is the

⁹An exception could be when the counterparty is liable to run into liquidity problems. Although the assets being hedged have increased in value, the counterparty might be unable to post collateral when required. An example here is Ashanti Goldfields (see Business Snapshot 21.2).

¹⁰See J. Hull and A. White, "CVA and Wrong Way Risk," *Financial Analysts Journal* 68, no. 5 (September/October 2012): 58–69.

¹¹DVA is sometimes also referred to as "debt value adjustment."

expected cost to the counterparty because the dealer might default. It is the counterparty's CVA. If DVA is a cost to the counterparty, it must be a benefit to the dealer. (This is because derivatives are zero sum games: the gain to one side always equals the loss to the other side.) The benefit arises from the fact that the dealer when it defaults avoids payments that would otherwise be required on outstanding derivatives. Accounting standards recognize both CVA and DVA. The book value of the derivatives outstanding with a counterparty to be calculated is

$$f_{\text{nd}} - \text{CVA} + \text{DVA}$$

where f_{nd} is the value of the derivatives assuming neither side will default.

DVA can be calculated at the same time as CVA. Equation (18.1) gives DVA rather than CVA if R is the recovery rate of the counterparty in the event of a default by the dealer, v_i is the present value of the counterparty's exposure to the dealer at the midpoint of the i th interval, and q_i is the probability of a default by the dealer during the i th interval. The counterparty's net exposure to the dealer takes account of collateral posted by the dealer with the counterparty in the same way that the dealer's net exposure to the counterparty takes account of the collateral posted by the counterparty with the dealer.

One surprising effect of DVA is that when the credit spread of a derivatives dealer increases, DVA increases. This in turn leads to an increase in the reported value of the derivatives on the books of the dealer and a corresponding increase in its profits. Some banks reported several billion dollars of profits from this source in the third quarter of 2011. Regulators are uncomfortable with this and have excluded DVA gains and losses from the definition of common equity in the determination of regulatory capital.

18.7 Some Simple Examples

To illustrate the ideas presented so far in this chapter, we now present some simple examples.

18.7.1 Single Transaction with Positive Value

Suppose first that a dealer has a single derivatives transaction with a counterparty that is bound to have a positive value to the dealer and a negative value to the counterparty at all future times. (An example would be the situation where the dealer has bought an option from the counterparty.) We suppose that no collateral has to be posted by the counterparty and, for ease of exposition, assume that payoffs on the derivatives transaction occur only on its expiration date.

The exposure of the dealer to the counterparty at a future time is the value of the transaction at that time. The present value of the expected exposure at time t_i , which we have denoted by v_i , is therefore the present value of the expected value of the transaction at time t_i . Because we are assuming no payoffs before maturity, the present value of the expected value of the transaction at time t_i is always equal to its value today.

Equation (18.1) therefore becomes

$$\text{CVA} = (1 - R)f_0 \sum_{i=1}^n q_i$$

where f_0 is the value of the derivative today, assuming no defaults. If f_0^* is the value after defaults are taken into account:

$$f_0^* = f_0 - \text{CVA}$$

or

$$f_0^* = f_0 \left[1 - (1 - R) \sum_{i=1}^n q_i \right] \quad (18.6)$$

This means that the effect of defaults is to reduce the value of the derivative by a proportional amount equal to the cumulative risk-neutral probability of default during the life of the derivative times one minus the recovery rate.

Now consider an unsecured zero-coupon bond issued by the counterparty that promises \$1,000 at time T . Define B_0 as the value of the bond assuming no possibility of default and B_0^* as the actual value of the bond. Assuming that the bond ranks equally with the derivative in the event of a default, it will have the same recovery rate. Similarly to equation (18.6)

$$B_0^* = B_0 \left[1 - (1 - R) \sum_{i=1}^n q_i \right] \quad (18.7)$$

From equations (18.6) and (18.7)

$$\frac{f_0^*}{f_0} = \frac{B_0^*}{B_0} \quad (18.8)$$

If γ is the yield on a risk-free zero-coupon bond maturing at time T and γ^* is the yield on a zero-coupon bond issued by the counterparty that matures at time T , then $B_0 = e^{-\gamma T}$ and $B_0^* = e^{-\gamma^* T}$ so that equation (18.8) gives

$$f_0^* = f_0 e^{-(\gamma^* - \gamma)T} \quad (18.9)$$

This shows that the derivative can be valued by increasing the discount rate that is applied to the expected payoff by $\gamma^* - \gamma$.

Example 18.2

Consider a two-year uncollateralized over-the-counter option sold by a company with a value, assuming no possibility of default, of \$3. Suppose that two-year zero-coupon bonds issued by the company have a yield that is 1.5% greater than a similar risk-free zero-coupon bond. The value of the option is

$$3e^{-0.015 \times 2} = 2.91$$

or \$2.91.

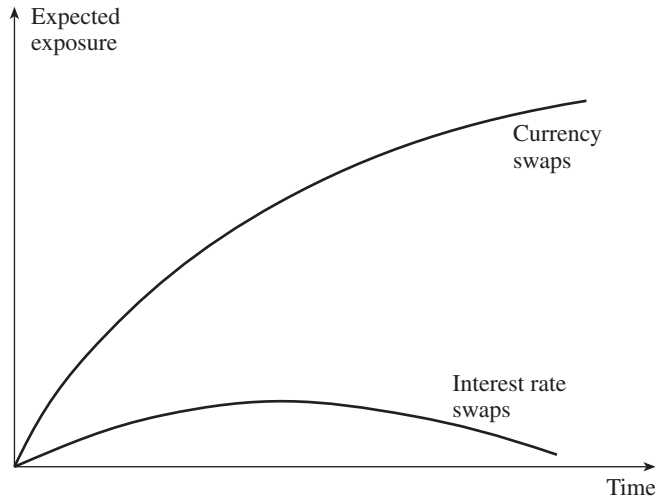


Figure 18.1 Expected Exposure on a Matched Pair of Offsetting Interest Rate Swaps and a Matched Pair of Offsetting Currency Swaps

18.7.2 Interest Rate Swaps vs. Currency Swaps

Consider next the situation where the dealer has entered into a pair of swaps that offset each other with two different counterparties. Figure 18.1 compares the total expected future exposure on the two transactions when they are (a) currency swaps and (b) interest rate swaps. The expected exposure on the interest rate swaps starts at zero, increases, and then decreases. By contrast, expected exposure on the currency swaps increases steadily with the passage of time. The main reason for the difference is that principals are exchanged at the end of the life of a currency swap and there is uncertainty about the exchange rate at that time. By contrast, toward the end of the life of the interest rate swap, very little is still to be exchanged. The impact of default risk for a dealer in currency swaps is therefore much greater than for a dealer in interest rate swaps. The q_i that the dealer calculates for counterparties in equation (18.1) are the same regardless of the transaction, but the v_i are on average greater for currency swaps.

18.7.3 Single Forward Transaction

For another example, we suppose that a dealer has a single forward transaction with a counterparty giving the dealer the right to buy an asset from the counterparty at time T for a price K . No collateral has to be posted. Suppose that the forward price of the asset today is F_0 (which is known) and at time t ($t \leq T$) it is F_t (which is unknown). As explained in Appendix C, the value of the transaction at time t is

$$(F_t - K)e^{-r(T-t)}$$

where r is the risk-free interest rate (assumed constant).

The exposure at time t is

$$\max[(F_t - K)e^{-r(T-t)}, 0] = e^{-r(T-t)}\max[(F_t - K), 0] \quad (18.10)$$

The present value of the exposure is therefore $e^{-r(T-t)}$ times the value of a derivative that pays off $\max[(F_t - K), 0]$ at time t . This is an option on a forward price. From Appendix E, the value of the derivative is

$$e^{-rt}[F_0N(d_1) - KN(d_2)]$$

where

$$d_1 = \frac{\ln(F_0/K) + \sigma^2 t/2}{\sigma\sqrt{t}}$$

and

$$d_2 = \frac{\ln(F_0/K) - \sigma^2 t/2}{\sigma\sqrt{t}}$$

and σ is the volatility of the forward price of the asset. It follows from equation (18.10) that the present value of the exposure at time t is

$$e^{-rT}[F_0N(d_1) - KN(d_2)]$$

The exposure, v_i , at a possible default time, t_i , is therefore

$$v_i = e^{-rt_i}[F_0N(d_{1,i}) - KN(d_{2,i})] \quad (18.11)$$

where

$$d_{1,i} = \frac{\ln(F_0/K) + \sigma^2 t_i/2}{\sigma\sqrt{t_i}}$$

and

$$d_{2,i} = \frac{\ln(F_0/K) - \sigma^2 t_i/2}{\sigma\sqrt{t_i}}$$

Example 18.3

Suppose that a bank has entered into a forward contract to buy 1 million ounces of gold from a mining company in two years for \$1,500 per ounce. The current forward price for the contract is \$1,600 per ounce. The probability of the company defaulting during the first year is 2% and the probability of the company defaulting during the second year is 3%. Defaults are assumed to happen at the midpoints of the years. The risk-free rate is 5% per annum. The financial institution anticipates a 30% recovery in the event of a default. The volatility of the forward price of gold when the forward contract expires in two years is 20%.

In this case, from equation (18.11)

$$\nu_1 = e^{-0.05 \times 2} [1,600N(d_{1,1}) - 1,500N(d_{2,1})]$$

where

$$d_{1,1} = \frac{\ln(1,600/1,500) + 0.2^2 \times 0.5/2}{0.2\sqrt{0.5}} = 0.5271$$

$$d_{2,1} = \frac{\ln(1,600/1,500) - 0.2^2 \times 0.5/2}{0.2\sqrt{0.5}} = 0.3856$$

so that $\nu_1 = 132.38$. Also, from equation (18.11)

$$\nu_2 = e^{-0.05 \times 2} [1,600N(d_{1,2}) - 1,500N(d_{2,2})]$$

where

$$d_{1,2} = \frac{\ln(1,600/1,500) + 0.2^2 \times 1.5/2}{0.2\sqrt{1.5}} = 0.3860$$

$$d_{2,2} = \frac{\ln(1,600/1,500) - 0.2^2 \times 1.5/2}{0.2\sqrt{1.5}} = 0.1410$$

so that $\nu_2 = 186.65$.

Other variables are: $q_1 = 0.02$, $q_2 = 0.03$, and $R = 0.3$ so that

$$\text{CVA} = (1 - 0.3) \times (0.02 \times 132.38 + 0.03 \times 186.65) = 5.77$$

The no-default value of the forward contract is $(1,600 - 1,500)e^{-2 \times 0.05} = 90.48$ or \$90.48. The value after the possibility of defaults has been taken into account is therefore

$$90.48 - 5.77 = 84.71$$

or \$84.71.

This example can be extended to the situation where the mining company can default more frequently or to calculate DVA (see Problems 18.13 and 18.14).

18.8 Other xVAs

Consider the situation in Figure 18.2. Bank A has entered into an uncollateralized interest rate swap transaction with a corporate end user and an offsetting transaction with another bank, Bank B. The latter is cleared through a CCP and collateralized with initial and variation margin. It appears that Bank A has locked in a profit of 0.1% because it receives 3% from Bank B and pays 2.9% to the corporate end user. However, the transaction with Bank B is liable to lead to incremental initial margin, which must be funded. The interest paid by the CCP on the initial margin is likely to be less than the cost of funding it. The excess of the funding cost over the interest received from the CCP gives rise to a margin value adjustment (MVA) and reduces the value of the transactions to Bank A.

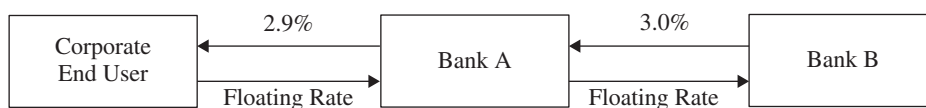


Figure 18.2 Swaps Entered into by Bank A; The Transaction with Bank B Is Cleared through a CCP

The transaction with the end user is cleared bilaterally with no collateral being posted.

The variation margin on Bank A's transaction with Bank B is not offset by the other transaction with the corporate end user because the latter is uncollateralized. When variation margin has to be paid to Bank B it has to be funded and, assuming that the funding cost is greater than the interest received from the CCP on variation margin, there is a cost; when variation margin is received from Bank B and the interest paid to the CCP is less than Bank A's normal funding cost, it is a source of funding. The net funding cost associated with variation margin is referred to as funding value adjustment (FVA). Because variation margin can be a use of funding or a source of funding, FVA can be positive or negative.

The capital valuation adjustment (KVA) is a charge to a derivatives transaction for the incremental capital requirements that the derivative gives rise to. In theory, equity capital is not a cost to a financial institution. In practice, financial institutions tend to regard equity capital as a cost because the expected return required by shareholders is higher than that required by debtholders.¹²

Summary

A derivatives dealer's credit value adjustment (CVA) for its bilaterally cleared derivatives with a counterparty is the present value of the expected future loss from the possibility of the counterparty defaulting. A derivatives dealer typically has systems in place for calculating the credit value adjustments (CVAs) for all the counterparties with which it has bilaterally cleared derivatives. The systems must simulate the market variables underlying outstanding transactions with each counterparty so that the expected net exposure at future times conditional on a default can be estimated. The calculations take account of netting and collateral agreements. The simplest assumption is that the probability of default is independent of the exposure. Procedures can be developed to incorporate wrong-way risk (where there is positive dependence between the two) and right-way risk (where there is negative dependence between the two).

CVAs are complex derivatives, and many dealers hedge them in the same way that they hedge other derivatives. There are two types of risks associated with CVAs: the risk

¹²Financial economists would argue that increasing equity capital makes a financial institution less risky and therefore means that all equity holders reduce the expected returns they require. This is not usually taken into account by financial institutions.

that the counterparty's credit spread will change and the risk that there will be movements in the values of the underlying market variables. Once a CVA has been calculated, it is relatively easy to quantify the first risk, and regulators require dealers to keep market risk capital for this risk. Calculating the sensitivity of CVA to movements in the underlying market variables is more difficult, but regulatory capital has been proposed for this risk as well.

A dealer's debit value adjustment (DVA) for its bilaterally cleared derivatives with a counterparty is the present value of the expected future loss to the counterparty from the possibility of the dealer defaulting. This is a gain to the dealer. It can be calculated at the same time as CVA. Other xVAs discussed in this chapter are margin value adjustment (MVA), funding value adjustment (FVA), and capital value adjustment (KVA).

Further Reading

- Basel Committee on Banking Supervision. "Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems," www.bis.org/publ/bcbs189dec1810.pdf, December 2010.
- Canabarro, E., and D. Duffie. "Measuring and Marking Counterparty Risk." Chapter 9 in *Asset/Liability Management for Financial Institutions*, edited by L. Tilman. New York: Institutional Investor Books, 2003.
- Gregory, J. *Counterparty Credit Risk: The New Challenge for Financial Markets*. 2nd ed. Hoboken, NJ: John Wiley & Sons, 2012.
- Hull, J., and A. White. "The Impact of Default Risk on the Prices of Options and Other Derivative Securities." *Journal of Banking and Finance* 19 (1995): 299–322.
- Pengelly, M. "CVA Melee." *Risk* 24, no. 2 (2011): 37–39.
- Picault, E. "Calculating and Hedging Exposure, CVA, and Economic Capital for Counterparty Credit Risk." In *Counterparty Credit Risk Modeling*, edited by M. Pykhtin. London: Risk Books, 2005.
- Sokol, A. "A Practical Guide to Monte Carlo CVA." Chapter 14 in *Lessons from the Crisis*, edited by A. Berd. London: Risk Books, 2010.

Practice Questions and Problems (Answers at End of Book)

- 18.1 Explain why a new transaction by a bank with a counterparty can have the effect of increasing or reducing the bank's credit exposure to the counterparty.
- 18.2 A company offers to post its own equity as collateral. How would you respond?
- 18.3 Suppose that a financial institution has two derivatives transactions outstanding with different counterparties, X and Y. Which of the following is/are true?
 - (a) The total expected exposure in one year on the two transactions is the sum of the expected exposure on the transaction with X and the expected exposure on the transaction with Y.
 - (b) The total present value of the cost of defaults is the sum of the present value of the cost of defaults on the transaction with X plus the present value of the cost of defaults on the transaction with Y.

- (c) The 95th percentile for the total exposure in one year on both transactions is the sum of the 95th percentile for the exposure in one year on the transaction with X and the 95th percentile for the exposure in one year on the transaction with Y.

Explain your answers.

- 18.4 “In the absence of collateral and other transactions between the parties, a long forward contract subject to credit risk is a combination of a short position in a no-default put and a long position in a call subject to credit risk.” Explain this statement.
- 18.5 Suppose that the spread between the yield on a three-year riskless zero-coupon bond and a three-year zero-coupon bond issued by a corporation is 120 basis points. By how much do standard option pricing models such as Black–Scholes–Merton overstate the value of a three-year option sold by the corporation? Assume there is only this one transaction between the corporation and its counterparty and no collateral is posted.
- 18.6 Can the existence of default triggers increase default risk? Explain your answer.
- 18.7 Give two examples of when (a) wrong-way risk and (b) right-way risk can be expected to be observed.
- 18.8 Explain the term “cure period.”
- 18.9 “Netting means that CVA cannot be calculated on a transaction-by-transaction basis.” Explain this statement.
- 18.10 “DVA can improve the bottom line when a bank is experiencing financial difficulties.” Explain why this is so.
- 18.11 What part of CVA risk is considered a component of market risk by regulators?
- 18.12 A CSA between a dealer and one of its counterparties states that collateral has to be posted by both sides with zero thresholds. If the cure period is assumed to be 10 days, under what circumstances will the dealer’s CVA model lead to losses?

Further Questions

- 18.13 Extend Example 18.3 to calculate CVA when default can happen in the middle of each month. Assume that the default probability during the first year is 0.001667 per month and the default probability during the second year is 0.0025 per month.
- 18.14 Calculate DVA for the bank in Example 18.3. Assume that the bank can default in the middle of each month and that the default probability is 0.001 per month for the two years. Assume that the recovery rate for the counterparty when the bank defaults is 40%.
- 18.15 Consider a European call option on a non-dividend-paying stock where the stock price is \$52, the strike price is \$50, the risk-free rate is 5%, the volatility is 30%, and the time to maturity is one year. Answer the following questions assuming no recovery in the event of default, that the probability of default is independent

of the option valuation, no collateral is posted, and no other transactions between the parties are outstanding.

- (a) What is the value of the option, assuming no possibility of a default?
- (b) What is the value of the option to the buyer if there is a 2% chance that the option seller will default at maturity?
- (c) Suppose that, instead of paying the option price up front, the option buyer agrees to pay the option price (with accumulated interest) at the end of the option's life. By how much does this reduce the cost of defaults to the option buyer in the case where there is a 2% chance of the option seller defaulting?
- (d) If in case (c) the option buyer has a 1% chance of defaulting at the end of the life of the option, what is the default risk to the option seller? Discuss the two-sided nature of default risk in the case and the value of the option to each side.

- 18.16 Suppose that the spread between the yield on a three-year riskless zero-coupon bond and a three-year zero-coupon bond issued by a bank is 210 basis points. The Black–Scholes–Merton price of an option is \$4.10. How much should you be prepared to pay for it if you buy it from a bank?

Chapter 19

Credit Value at Risk

The value at risk (VaR) measure is central to the determination of regulatory capital for credit risk and to much of the credit risk management carried out internally by both financial and nonfinancial corporations. This chapter covers alternative approaches for calculating credit risk VaR.

Credit risk VaR is defined similarly to market risk VaR. It is the credit risk loss over a certain time period that will not be exceeded with a certain confidence level. Some credit risk VaR models consider only losses from defaults; others consider losses from downgrades or credit spread changes as well as from defaults.

Banks calculate credit risk VaR to determine both regulatory capital and economic capital. The regulatory capital requirements for credit risk are discussed in Chapters 25 and 26. Economic capital, which is discussed in Chapter 28, is a financial institution's own estimate of the capital it requires for the risks it is taking and is used to calculate return on capital measures for its business units. Sometimes the VaR model that a bank chooses to use to determine credit risk economic capital is different from the one it is required to use for the determination of regulatory capital.

Value at risk is a scenario analysis measure. As explained in Chapter 10, the probability of default estimates used to develop alternative scenarios should therefore be real-world estimates. Risk-neutral estimates should then be used to value the underlying

portfolio on the VaR horizon date. It will be recalled from Chapter 17 that real-world default probabilities are lower than risk-neutral default probabilities.

The time horizon for credit risk VaR is often longer than that for market risk VaR. Market risk VaR is often calculated with a one-day time horizon. Credit risk VaR, for instruments that are not held for trading, is usually calculated with a one-year time horizon. Historical simulation (see Chapter 12) is the main tool used to calculate market risk VaR, but a more elaborate model is usually necessary to calculate credit risk VaR.

A key aspect of any credit risk VaR model is credit correlation. Defaults for different companies do not happen independently of each other. During an economic downturn, most companies are adversely affected and become more likely to default. When the economy is faring well, they are favorably affected and less likely to default. This relationship between default rates and economic factors is a major reason for credit correlation. If credit correlation increases (as it tends to do in stressed economic conditions), the risk for a financial institution with a portfolio of credit exposures increases.

19.1 Ratings Transition Matrices

The methods used by financial institutions for calculating credit VaR often involve ratings transition matrices. These are matrices showing the probability of a company migrating from one rating category to another during a certain period of time. They are based on historical data. The rating categories can be either those used internally by the financial institution or those produced by rating agencies such as Moody's, S&P, and Fitch. Table 19.1 shows the one-year transition matrix produced by S&P, based on following the performance of all the companies that S&P rated between 1981 and 2020. For example, it shows that a company that starts with a rating of A has a 92.61% probability of still being rated A at the end of a year. There is a 1.67% chance that it will be upgraded to AA by the end of the year, a 5.23% chance that it will be downgraded to BBB, and so on. The probability that it will default during the year is only 0.05%, or 5 chances in 10,000.

If we assume that the rating change in one period is independent of that in another period, Table 19.1 can be used to calculate a transition matrix for periods other than one year. For example, a transition matrix for two years can be calculated by multiplying the matrix by itself. The five-year transition matrix, which is shown in Table 19.2, is calculated as the fifth power of the matrix in Table 19.1. Table 19.2 shows, not surprisingly, that the probability of a company keeping the same credit rating over five years is much lower than it is over one year, and default probabilities over five years are much higher than over one year.

The credit rating change over a period less than a year is not quite so easy to calculate. For example, estimating a transition matrix for six months involves taking the square root of the matrix in Table 19.1; estimating the transition matrix for three months involves taking the fourth root of the matrix; and so on. The calculation methodology is explained

Table 19.1 One-Year Ratings Transition Matrix, 1981–2020, with Probabilities Expressed as Percentages and Adjustments for Transition to the NR (not rated) Category

Initial Rating	Rating at Year-end							
	AAA	AA	A	BBB	BB	B	CCC/C	D
AAA	89.85	9.35	0.55	0.05	0.11	0.03	0.05	0.00
AA	0.50	90.76	8.08	0.49	0.05	0.06	0.02	0.02
A	0.03	1.67	92.61	5.23	0.27	0.12	0.02	0.05
BBB	0.00	0.10	3.45	91.93	3.78	0.46	0.11	0.17
BB	0.01	0.03	0.12	5.03	86.00	7.51	0.61	0.70
B	0.00	0.02	0.08	0.17	5.18	85.09	5.66	3.81
CCC/C	0.00	0.00	0.12	0.20	0.65	14.72	50.90	33.42
D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00

SOURCE: S&P.

Table 19.2 Five-Year Ratings Transition Matrix Calculated from Table 19.1 with Probabilities Expressed as Percentages

Initial Rating	Rating after 5 years							
	AAA	AA	A	BBB	BB	B	CCC/C	D
AAA	58.92	31.31	7.67	1.12	0.47	0.25	0.10	0.15
AA	1.69	62.98	28.97	5.04	0.62	0.37	0.07	0.22
A	0.17	6.03	70.62	19.34	2.41	0.84	0.13	0.47
BBB	0.02	0.80	12.77	68.51	12.35	3.53	0.50	1.53
BB	0.04	0.20	1.76	16.24	50.90	21.37	2.64	6.91
B	0.01	0.10	0.47	2.42	14.56	50.61	7.31	24.59
CCC/C	0.00	0.03	0.31	0.86	4.14	18.88	5.62	70.18
D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00

in Appendix J at the end of the book and software for performing the calculations is on the author's website: www-2.rotman.utoronto.ca/~hull/riskman.

Table 19.3 shows the rating transition calculated for a period of one month from the data in Table 19.1. As might be expected, the probability that a company's credit rating will stay the same over a period of one month is very high.

The assumption that the credit rating change in one period is independent of that in another period is not exactly true. If a company has been downgraded recently, it is more likely to be downgraded again in the next short period of time.¹ (This phenomenon is

¹For a discussion of this, see E. Altman and D. Kao, "The Implications of Corporate Bond Rating Drift," *Financial Analysts Journal* (May–June 1992): 64–75, and D. Lando and T. Skodeberg, "Analyzing Rating Transitions and Rating Drift with Continuous Observations," *Journal of Banking and Finance* 26 (2002): 423–444.

Table 19.3 One-Month Ratings Transition Matrix Calculated from Table 19.1 with Probabilities Expressed as Percentages

Initial Rating	Rating after 1 month							
	AAA	AA	A	BBB	BB	B	CCC/C	D
AAA	99.11	0.86	0.02	0.00	0.01	0.00	0.01	0.00
AA	0.05	99.19	0.73	0.03	0.00	0.01	0.00	0.00
A	0.00	0.15	99.35	0.47	0.02	0.01	0.00	0.00
BBB	0.00	0.01	0.31	99.28	0.35	0.03	0.01	0.01
BB	0.00	0.00	0.00	0.47	98.72	0.72	0.04	0.04
B	0.00	0.00	0.01	0.00	0.50	98.58	0.69	0.22
CCC/C	0.00	0.00	0.01	0.02	0.02	1.79	94.45	3.71
D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00

sometimes referred to as ratings momentum.) However, the independence assumption is not too unreasonable for some purposes.²

19.2 Vasicek's Model

Vasicek's Gaussian copula model (see Section 9.6) is a way of calculating high percentiles of the distribution of the default rate for a portfolio of loans. As in Chapter 9, we define $\text{WCDR}(T, X)$ as the X th percentile of the default rate distribution during a period of length T . (It will be recalled that WCDR denotes worst case default rate.) Vasicek's model relates $\text{WCDR}(T, X)$ to the probability of default in time T , PD, and a parameter, ρ , describing credit correlation. The formula, which is proved in Section 9.6, is

$$\text{WCDR}(T, X) = N\left(\frac{N^{-1}(\text{PD}) + \sqrt{\rho}N^{-1}(X)}{\sqrt{1 - \rho}}\right) \quad (19.1)$$

For an individual loan, if EAD is the exposure at default and LGD is the loss given default, the X th percentile of the loss distribution is

$$\text{WCDR}(T, X) \times \text{EAD} \times \text{LGD}$$

A result by Gordy (2003) enables us to extend this in a one-factor world.³ If we have a large portfolio of n loans where each loan is a small part of the total portfolio, the X th percentile of the loss distribution is approximately

²When the five-year transition matrix in Table 19.2 is compared with the actual five-year transition matrix, there are some differences. For example, companies in the top three rating categories are more likely to default or suffer big downgrades. This is consistent with ratings momentum.

³See M. Gordy, "A Risk-Factor Model Foundation for Ratings-Based Capital Rules," *Journal of Financial Intermediation* 12, no. 3 (July 2003): 199–233.

$$\sum_{i=1}^n \text{WCDR}_i(T, X) \times \text{EAD}_i \times \text{LGD}_i \quad (19.2)$$

where $\text{WCDR}_i(T, X)$, EAD_i , and LGD_i are the values of WCDR, EAD, and LGD for the i th loan in the portfolio.

As will be explained in Chapter 25, regulatory capital for the banking book is set equal to the expression in equation (19.2) with T equal to one year and X equal to 99.9%. Sometimes the expression under the summation sign in equation (19.2) is multiplied by a maturity adjustment factor, to allow for the fact that, if it lasts longer than one year, the i th loan might deteriorate without defaulting.

Structural models of the type discussed in Section 17.8 can be used to show that ρ for two companies should be roughly equal to the correlation between the returns on the assets of the companies.⁴ As an approximation, this is the same as the correlation between the returns on their equities. One way of determining ρ for a portfolio of exposures to companies is therefore to calculate the average correlation between the returns on equities for the companies. If the companies are not publicly traded, the average correlation between similar publicly traded companies can be used instead.

As pointed out in Chapter 9, Vasicek's model has the disadvantage that it incorporates very little tail correlation. Replacing the Gaussian copula model with another copula model (see Section 9.6.3 for some alternatives) can remedy this.

19.3 Credit Risk Plus

In 1997, Credit Suisse Financial Products proposed a methodology for calculating VaR that it termed Credit Risk Plus.⁵ It involves analytic approximations that are well established in the insurance industry.

Suppose that a financial institution has n loans of a certain type and the probability of default for each loan during a year is q so that the expected number of defaults is qn . Assuming that default events are independent, the probability of m defaults is given by the binomial distribution as

$$\frac{n!}{m!(n-m)!} q^m (1-q)^{n-m}$$

If q is small and n large, this can be approximated by the Poisson distribution:

$$\text{Prob}(m \text{ defaults}) = \frac{e^{-qn} (qn)^m}{m!} \quad (19.3)$$

This equation is approximately true even if the probability of default is not the same for each loan, provided that all the probabilities of default are small and q equals the average probability of default over the next year for the loans in the portfolio.

⁴See, for example, J. Hull, M. Predescu, and A. White, "The Valuation of Correlation Dependent Derivatives Using a Structural Model," *Journal of Credit Risk* 6, no. 3 (Fall 2010): 99–132.

⁵See Credit Suisse Financial Products, "Credit Risk Management Framework" (October 1997).

In practice, we are uncertain about what the default rate, q , will be during the next year. As shown in Table 9.6, default rates vary greatly from year to year. A convenient assumption is that qn , the expected number of defaults, has a gamma distribution with mean μ and standard deviation σ .

The Poisson distribution in equation (19.3) then becomes a negative binomial distribution

$$\text{Prob}(m \text{ defaults}) = p^m(1-p)^\alpha \frac{\Gamma(m+\alpha)}{\Gamma(m+1)\Gamma(\alpha)} \quad (19.4)$$

where $\alpha = \mu^2/\sigma^2$, $p = \sigma^2/(\mu + \sigma^2)$, and $\Gamma(x)$ is the gamma function.⁶

Table 19.4 shows the probability distribution for the number of defaults when $\mu = 4$ for various values of σ . The first column ($\sigma = 0$) is calculated from equation (19.3). The remaining columns are calculated from equation (19.4). It can be seen that as σ tends to zero, the negative binomial distribution in equation (19.4) tends to the same probability distribution for the number of defaults as the Poisson distribution. As σ increases, the probability of an extreme outcome involving a large number of defaults increases. Suppose that the loss from a default is constant. As indicated in the final row of the table, when $\sigma = 0$ so that there is no uncertainty about the default rate, the 99.9% VaR is calculated as the loss when there are 11 defaults. When $\sigma = 10$ so that there is a great deal of uncertainty about the default rate, the 99.9% VaR is considerably larger. It is the loss when there are 98 defaults.

A more flexible way of building uncertainty into the default rate is by using Monte Carlo simulation. Typically there are several different categories of exposures with different but correlated default rates. One way of proceeding is as follows:

1. Sample an overall default rate. This could be a random sample from data such as that in Table 9.6.
2. Develop a model relating the default rate for each category of obligors to the overall default rate. A simple model is obtained by regressing default rates for each category of obligors against the overall default rate.
3. Sample a number of defaults for each category of obligors.
4. Sample a loss given default for each default in each category.
5. Calculate the total loss from the defaults.
6. Repeat steps 1 to 5 many times to build up a probability distribution for the total loss.
7. Calculate the required VaR from the total loss probability distribution.

⁶In Excel, the GAMMALN function returns the natural logarithm of the gamma function, $\Gamma(x)$. When x is an integer $\Gamma(x) = (x-1)!$

Table 19.4 Probability Distribution for Number of Defaults When the Expected Number of Defaults Is 4.0.

Number of Defaults	Standard Deviation of Default Rate						
	0	0.1	0.5	1.0	2.0	5.0	10.0
0	0.0183	0.0184	0.0207	0.0281	0.0625	0.2814	0.5938
1	0.0733	0.0734	0.0777	0.0901	0.1250	0.1553	0.0913
2	0.1465	0.1466	0.1486	0.1531	0.1563	0.1098	0.0509
3	0.1954	0.1952	0.1923	0.1837	0.1563	0.0833	0.0353
4	0.1954	0.1951	0.1895	0.1746	0.1367	0.0653	0.0268
5	0.1563	0.1561	0.1516	0.1396	0.1094	0.0523	0.0214
6	0.1042	0.1041	0.1026	0.0978	0.0820	0.0423	0.0177
7	0.0595	0.0596	0.0603	0.0614	0.0586	0.0346	0.0150
8	0.0298	0.0298	0.0315	0.0353	0.0403	0.0285	0.0129
9	0.0132	0.0133	0.0148	0.0188	0.0269	0.0236	0.0113
10	0.0053	0.0053	0.0064	0.0094	0.0175	0.0196	0.0099
11	0.0019	0.0019	0.0025	0.0045	0.0111	0.0163	0.0088
12	0.0006	0.0007	0.0009	0.0020	0.0069	0.0137	0.0079
13	0.0002	0.0002	0.0003	0.0009	0.0043	0.0115	0.0071
14	0.0001	0.0001	0.0001	0.0004	0.0026	0.0096	0.0064
15	0.0000	0.0000	0.0000	0.0001	0.0016	0.0081	0.0058
16	0.0000	0.0000	0.0000	0.0001	0.0009	0.0068	0.0053
17	0.0000	0.0000	0.0000	0.0000	0.0005	0.0058	0.0048
18	0.0000	0.0000	0.0000	0.0000	0.0003	0.0049	0.0044
19	0.0000	0.0000	0.0000	0.0000	0.0002	0.0041	0.0041
20	0.0000	0.0000	0.0000	0.0000	0.0001	0.0035	0.0038
99.9 percentile of number of defaults	11	11	12	13	17	39	98

There are a number of ways in which the Monte Carlo simulation can be refined. An analysis of the data in Table 9.6 shows that the default rate in one year is not independent of the default rate in the previous year. This indicates that randomly sampling a default rate from a table such as Table 9.6 to determine next year's default rate may not be the best approach. It may be preferable to develop a model that relates the default rate in one year to the default rate in the previous year or other economic factors observable in the previous year.

Note that the uncertainty about the default rate plays a key role in the analysis. Without this uncertainty, there is no default correlation and, as indicated in Table 19.4, there is very little chance of a large number of defaults. As the uncertainty about the default rate increases, default correlation increases and a large number of defaults becomes more likely. The default correlation arises because all companies have the same default rate, which can be high or low. Without default correlation, the loss probability distribution is fairly symmetrical. With default correlation, it is positively skewed, as indicated in Figure 19.1.

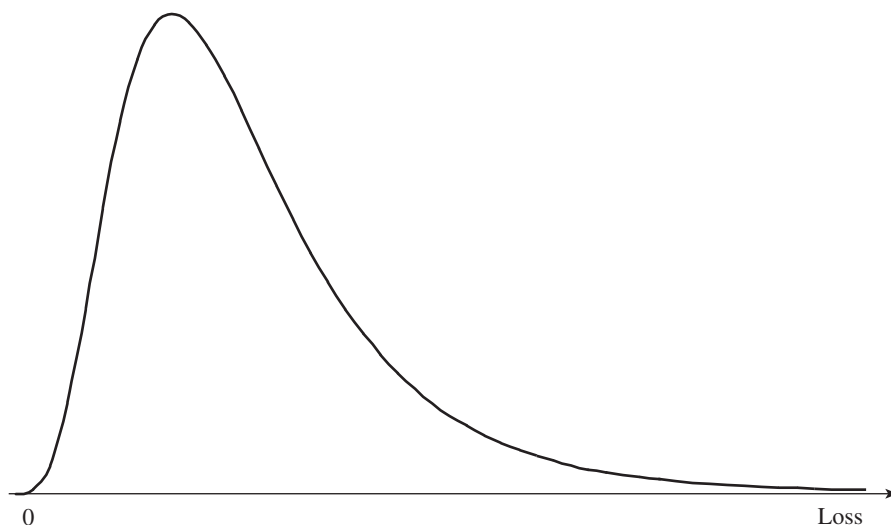


Figure 19.1 General Shape of Probability Distribution of Default Losses

19.4 Creditmetrics

Vasicek's model and Credit Risk Plus estimate the probability distribution of losses arising from defaults. The impact of downgrades is not considered.⁷ We now consider the CreditMetrics model, which is designed so that it can take account of downgrades as well as defaults. This model was proposed by JPMorgan in 1997. It is based on a rating transition matrix such as the one shown in Table 19.1. The ratings can be the internal ratings used by the bank (derived from its own historical data) or those produced by rating agencies.

Calculating a one-year credit VaR for a portfolio of transactions with many counterparties involves carrying out a Monte Carlo simulation. On each simulation trial, the credit ratings of all counterparties at the end of one year are determined. (The procedure for doing this will be explained shortly.) The credit loss for each counterparty is then calculated. If the end-of-year credit rating for a counterparty is not default, the calculation of the credit loss is given by valuing all the transactions with the counterparty at the one-year point. If the end-of-year credit rating is default, the credit loss is the exposure at default times one minus the recovery rate.⁸

The calculations require the term structure of credit spreads for each rating category. A simple assumption is that it is the same as that observed in the market today. An alternative assumption is that there is a credit spread index that has a certain probability distribution and that all credit spreads are linearly dependent on the index.

⁷However, the maturity adjustment factor does allow for downgrades in the regulatory implementation of Vasicek's model.

⁸The recovery rate is typically sampled from an empirical distribution.

When transactions with a counterparty involve derivatives, an extension of the CVA calculations described in Chapter 18 is necessary. Recall that, for each counterparty, a CVA calculation involves dividing the time between today and the end of the life of the longest transaction with the counterparty into n intervals. We defined q_i as the probability of default during the i th time interval, v_i is the present value of the expected net exposure at the midpoint of the interval after collateral has been taken into account, and R is the recovery rate. The expected cost of defaults from one year onward, as seen at time zero, is

$$\sum_{i=j}^n (1 - R)q_i v_i$$

where the j th interval starts at time one year.

On each CreditMetrics simulation trial where there is no default during the first year, the term structure of credit spreads at the one-year point gives the probability of default in each interval from year one onward using equation (18.3). Suppose that q_i^* is the probability of default for the i th interval ($i \geq j$) on a particular simulation trial. The credit loss for that trial is

$$\sum_{i=j}^n (1 - R)(q_i^* - q_i)v_i \quad (19.5)$$

Note that, if the counterparty's credit rating improves during the year, this credit loss is likely to be negative.

If there is a default during the first year on a particular simulation trial, the timing of the default is sampled and the exposure at default determined. This is multiplied by one minus the recovery rate to determine the loss.

The results of the CreditMetrics Monte Carlo simulation determine a probability distribution for total credit losses arising from defaults and downgrades across all counterparties. The required value at risk can be calculated from this distribution.

19.4.1 The Correlation Model

In sampling to determine credit losses, the credit rating changes for different counterparties are not assumed to be independent. A Gaussian copula model is used to construct a joint probability distribution of rating changes. (See Section 9.5 for a discussion of copula models.) The copula correlation between the rating transitions for two companies is typically set equal to the correlation between their equity returns. (The capital asset pricing model discussed in Chapter 1 provides a one-factor model for equity returns. As described in Section 9.4.2, this can be used to determine an internally consistent correlation matrix.)

Suppose that we are simulating the rating change of an A-rated and a B-rated company over a one-year period using the transition matrix in Table 19.1. Suppose that the correlation between the equity returns of the two companies is 0.2. On each simulation trial we would sample two variables x_A and x_B from standard normal distributions so

that their correlation is 0.2. The variable x_A determines the new rating of the A-rated company and variable x_B determines the new rating of the B-rated company. From Table 19.1, the probability of an A-rated company moving to AAA, AA, A, ... is 0.0003, 0.167, 0.9261 ... Because

$$\begin{aligned} N^{-1}(0.0003) &= -3.3416 \\ N^{-1}(0.0003 + 0.0167) &= -2.1201 \\ N^{-1}(0.0003 + 0.0167 + 0.9261) &= 1.5813 \end{aligned}$$

the A-rated company gets upgraded to AAA if $x_A < -3.3416$, it becomes AA-rated if $-3.3416 < x_A < -2.1201$, it stays A-rated if $-2.1201 < x_A < 1.5813$, and so on. Table 19.1 also shows that the probability of a B-rated company moving to AA, A, and BBB, ... is 0.0002, 0.0008, 0.0017, Because

$$\begin{aligned} N^{-1}(0.0002) &= -3.5401 \\ N^{-1}(0.0002 + 0.0008) &= -3.0902 \\ N^{-1}(0.0002 + 0.0008 + 0.0017) &= -2.7822 \end{aligned}$$

the B-rated becomes AA-rated if $x_B < -3.5401$, it becomes A-rated if $-3.5401 < x_B < -3.0902$, it becomes BBB-rated if $-3.0902 < x_B < -2.7822$, and so on. The A-rated company defaults if $x_A > N^{-1}(0.9995)$, that is, when $x_A > 3.2905$. The B-rated company defaults when $x_B > N^{-1}(0.9619)$, that is, when $x_B > 1.7744$. The full set of transitions are shown in Figure 19.2.

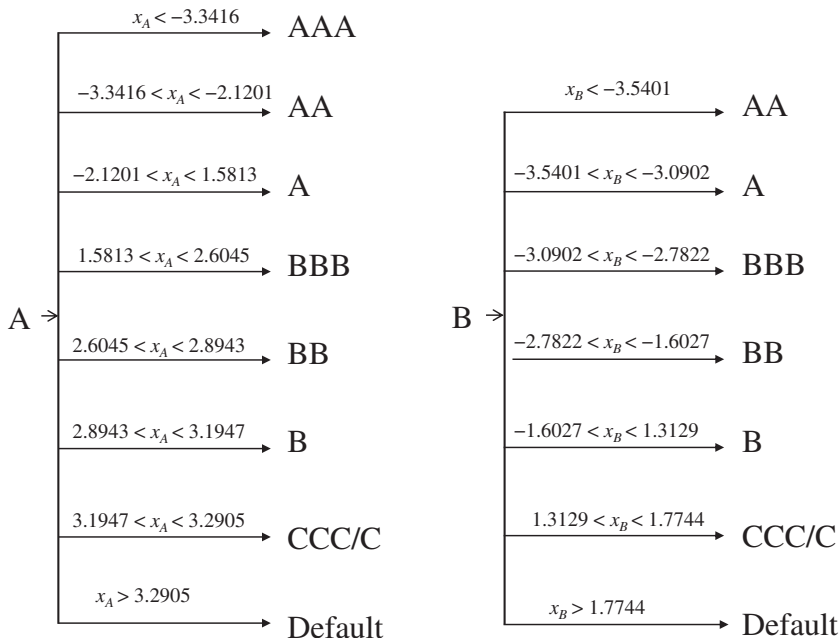


Figure 19.2 The CreditMetrics Correlation Model: Transition of A-Rated and B-Rated Companies to a New Rating after One Year

x_A and x_B are sampled from standard normal distributions with the correlation equal to the correlation between the equity returns of A and B.

It is interesting to note that, if both the CreditMetrics and the Credit Risk Plus models were based on the same set of assumptions, they should in theory predict the same probability distribution for losses over the long term. It is the timing of losses that is different. Suppose, for example, that you hold a certain loan in your portfolio. In year 1, the borrower gets downgraded from A to BBB; in year 2, it gets downgraded from BBB to B; in year 3 it defaults. You could assume that there are no losses in years 1 and 2 and calculate loss in year 3 (the Credit Risk Plus approach). Alternatively, you can calculate separate revaluation losses in years 1, 2, and 3 (the CreditMetrics approach). The losses under the second approach should in theory add up to the losses under the first approach.

19.5 Credit Spread Risk

The values of most credit-sensitive products in the trading book depend critically on credit spreads. Calculating VaR or ES for portfolios including these products therefore involves examining potential credit spread changes. One possibility is to use historical simulation to calculate a one-day 99% VaR. This one-day 99% VaR can then be scaled up by $\sqrt{10}$ to obtain a 10-day 99% VaR. It is first necessary to collect credit spread changes for each company to which the bank is exposed over the last 500 days (or some other period). The first historical simulation scenario assumes that the credit spread changes for all companies are the same as they were on the first day; the second scenario assumes that they are the same as they were on the second day; and so on. However, there are problems with this approach. If the company is alive today, it did not default in the past and the calculations carried out therefore assume no probability of default in the future.⁹ Another problem is that credit spreads are not updated daily for all companies and so the historical data may not be of high quality.

An alternative approach is to use a version of the CreditMetrics approach. A 10-day rating transitions matrix is calculated as indicated in Section 19.1. The ratings transition matrix defines the probability of a company moving from one rating category to another during a 10-day period. The historical data on credit rating changes define a probability distribution for the credit spread changes associated with each rating category at the end of 10 days. A Monte Carlo simulation is carried out. On each trial, the transition matrix is sampled from to determine whether each company stays in the same rating category, changes to a different rating category, or defaults. The probability distribution of credit spread changes is also sampled from to determine the credit spread at the end of the 10 days associated with each rating category. This enables the value of the portfolio at the end of 10 days to be calculated on each trial and a VaR to be obtained.

⁹One way of dealing with this is to assume that default takes place when the credit spread exceeds a certain level. Another is to separately consider “jump-to-default” risk, which is the approach taken by regulators.

Credit correlation can be introduced in two ways. First, a Gaussian copula model can be used to correlate the credit rating changes of different companies. Second, the credit spread changes for different rating categories can be assumed to be perfectly (or nearly perfectly) correlated so that when the spreads for A-rated instruments move up, those for instruments with other ratings do as well.

Example 19.1

Consider a company that owns a single two-year zero-coupon bond with a principal of \$1,000. Suppose that the risk-free rate is 3% and the current credit spread is 200 basis points so that the bond yield is 5%. (Rates are expressed with annual compounding.) The current price of the bond is $1,000/1.05^2 = \$907.03$. Suppose that the bond's current rating is BB and that during the next month there is a 0.6% chance that it will increase to BBB, a 98.5% chance that it will stay the same, a 0.8% chance that it will decrease to B, and a 0.1% chance that it will default. If it defaults, the bond will be worth \$400. For each possible rating category there are three equally likely credit spreads. In basis points these are 80, 100, and 120 for BBB; 160, 200, and 240 for BB; and 400, 450, and 500 for B.

The worst outcome is clearly a default. This has a probability of 0.1% and will lead to a loss of $907.03 - 400 = \$507.03$. The next worst outcome involves a transition to a rating of B when the credit spread for that rating category is 500 basis points. This has a probability of $0.8/3 = 0.267\%$. Because the bond will have about 1.917 years to maturity, its price will be $1,000/1.08^{1.917} = \$862.85$, for a loss of $907.03 - 862.85 = \$44.17$. A complete list of outcomes with their probabilities is shown in Table 19.5.

Table 19.5 Outcomes for Example 19.1

Rating	Spread (bp)	Probability	Bond Value (\$)	Loss (\$)
Default		0.100%	400.00	507.03
B	500	0.267%	862.85	44.17
B	450	0.267%	870.56	36.47
B	400	0.267%	878.38	28.65
BB	240	32.833%	904.11	2.92
BB	200	32.833%	910.72	-3.70
BB	160	32.833%	917.41	-10.38
BBB	120	0.200%	924.17	-17.14
BBB	100	0.200%	927.58	-20.55
BBB	80	0.200%	931.01	-23.98

The table shows that, if the confidence level is higher than 99.9%, the VaR is \$507.03; if it is between 99.9% and 99.633%, the VaR is \$44.17; and so on. When the confidence level is 99%, the VaR is a modest \$2.92.

19.5.1 Constant Level of Risk Assumptions

When calculating VaR or ES, banks sometimes make a constant level of risk assumption. This means that they assume a portfolio is rebalanced periodically to bring the risk back to where it was initially.

Suppose, for example, that a company invests in BBB bonds and is considering two alternative trading strategies:

1. Hold bonds for one year and then sell (the “buy and hold” strategy).
2. Rebalance at the end of each month so that if bonds are no longer BBB they are sold and replaced with BBB bonds (the “constant level of risk” strategy).

The first strategy leads to outcomes given by the BBB row of Table 19.1, while, in the case of the second strategy, the outcomes given by the BBB row of Table 19.3 are repeated 12 times.

Consider first defaults. From Table 19.1, the probability of a default under the buy-and-hold strategy is 0.0017. From Table 19.3, the probability of a default each month is 0.0001. The probability of one default when the constant level of risk strategy is used is therefore

$$12 \times 0.0001 \times 0.9999^{11} = 0.0012$$

This is less than that under the buy-and-hold strategy. The probability of two or more defaults is negligible.

Now, consider the probability of a downgrade to BB. Under the buy-and-hold strategy, the probability of this is 0.0378 from Table 19.1. When the constant level of risk strategy is used, the probability of one downgrade to BB is

$$12 \times 0.0035 \times 0.9965^{11} = 0.04404$$

which is greater than that for the buy-and-hold strategy. The probability of more than one downgrade is in this case about 0.001.

The calculations illustrate that the buy-and-hold strategy will lead to greater losses from defaults and big downgrades than the constant level of risk strategy. But the probabilities of losses from a small rating change are less. VaR and ES are generally smaller when the constant level of risk strategy is used.

Summary

Credit VaR is defined similarly to the way VaR is defined for market risk. It is the credit loss that will not be exceeded over some time horizon with a specified confidence level. Basel regulations require credit VaR to be calculated with a one-year time horizon and a 99.9% confidence level. For economic capital purposes a different confidence level may be used.

We have discussed three methods for calculating credit VaR: Vasicek's model, Credit Risk Plus, and CreditMetrics. Vasicek's model is based on a one-factor Gaussian copula model of the time to default for the companies in a portfolio. It is used in the determination of regulatory capital. Credit Risk Plus uses procedures that are similar to those used in the insurance industry to calculate default losses from assumptions about the probabilities of default of individual companies. CreditMetrics is different from the other two methods in that it considers losses from both defaults and downgrades. It uses a Gaussian copula model in conjunction with a ratings transition matrix.

To determine VaR or ES for instruments dependent on credit spreads, one approach is to collect historical data on the credit spread changes of companies and use a historical simulation methodology similar to that for market risk. Another is to model the rating transitions of companies and movements in the average credit spreads associated with different rating categories.

Further Reading

- Credit Suisse Financial Products. "Credit Risk Management Framework" (October 1997).
Finger, C. C. "Creditmetrics and Constant Level of Risk." Working Paper, RiskMetrics Group-MSCI, September 2010.
Gordy, M. "A Risk-Factor Model Foundation for Ratings-Based Capital Rules." *Journal of Financial Intermediation* 12, no. 3 (July 2003): 199–233.
JP Morgan. "CreditMetrics—Technical Document" (April 1997).
Vasicek, O. "Probability of Loss on a Loan Portfolio." Working Paper, KMV, 1987. (Published in *Risk* in December 2002 under the title "Loan Portfolio Value.")

Practice Questions and Problems (Answers at End of Book)

- 19.1 Explain the difference between Vasicek's model, the Credit Risk Plus model, and CreditMetrics as far as the following are concerned: (a) when a credit loss is recognized and (b) the way in which default correlation is modeled.
- 19.2 Explain what is meant by the constant level of risk assumption.
- 19.3 Use the transition matrix in Table 19.1 to calculate a transition matrix for two years. What is the probability of a company rated AAA being AAA at the end of two years? What is the probability of it moving to AA?
- 19.4 Use the transition matrix in Table 19.1 and software on the author's website to calculate a transition matrix for six months. What is the probability of a company rated Aaa staying Aaa during the six months? What is the probability of it moving to Aa?
- 19.5 How can historical simulation be used to calculate a one-day 99% VaR for the credit risk of bonds in the trading book? What are the disadvantages?
- 19.6 A bank has 100 one-year loans, each with a 1% probability of default. What is the probability of six or more defaults, assuming independence?

- 19.7 Repeat Problem 19.6 on the assumption that the probability of default applicable to all loans is uncertain. It is equally likely to be 0.5% or 1.5%.
- 19.8 What is the autocorrelation for the default rates in Table 9.6? What are the implications of this for a Credit Risk Plus model?

Further Questions

- 19.9 Explain carefully the distinction between real-world and risk-neutral default probabilities. Which is higher? A bank enters into a credit derivative where it agrees to pay \$100 at the end of one year if a certain company's credit rating falls from A to BBB or lower during the year. The one-year risk-free rate is 5%. Using Table 19.1, estimate a value for the derivative. What assumptions are you making? Do they tend to overstate or understate the value of the derivative?
- 19.10 Suppose that a bank has a total of \$10 million of small exposures of a certain type. The one-year probability of default is 1% and the recovery rate averages 40%. Estimate the 99.5% one-year credit VaR using Vasicek's model if the copula correlation parameter is 0.2.
- 19.11 Use the transition matrix in Table 19.1 and software on the author's website to calculate the transition matrix over 1.25 years.

Part Five

OTHER RISKS

Chapter 20

Operational Risk

In 1999, bank regulators announced plans to assign capital for operational risk. This met with some opposition from banks. The chairman and CEO of one major international bank described it as “the dopicst thing I have ever seen.” However, bank supervisors did not back down, because they knew that many of the big losses of banks over the previous 10 years had been operational risk losses—not credit risk or market risk losses.

The initial regulations gave banks a choice between two relatively simple ways of calculating operational risk capital (*the basic indicator approach* and the *standardized approach*) and one quite complicated approach (*the advanced measurement approach*, AMA). Regulators have since replaced these three approaches with a single “standardized measurement approach” that is simpler than AMA, but more complicated than the other two approaches. The new approach, which came into effect in 2023, is described in Section 20.4.

Operational risk has led to some big losses since the 1999 announcement (and this has probably helped to convince regulators that they were right to insist on requiring capital for operational risk). A big rogue trader loss at Société Générale in 2008 is discussed in Business Snapshot 5.5, and there was a similar loss at UBS in 2011. The London Whale loss at JPMorgan Chase in 2012 is discussed in Business Snapshot 22.1. In 2014, it was announced that the French bank BNP Paribas would pay \$9 billion (roughly one year’s profit) to the U.S. government for violating economic sanctions by moving dollar-denominated transactions through the American banking system on behalf of Sudanese, Iranian, and Cuban parties. The bank was also banned from conducting certain U.S. transactions for a year.

An increasingly important type of operational risk for banks is cyber risk. Banks have sophisticated systems in place to protect themselves from cyber attacks, but the attacks themselves are becoming more sophisticated. Also, banks are making increasing use of computer systems and the Internet, providing more avenues for cyber fraud. Customers and employees have to be continually educated so that the bank's data remain secure. Cyber attacks on banks are attractive to criminals because, to quote the bank robber Willie Sutton, "that's where the money is." They are also attractive to terrorists because of their potential to damage a nation's economic security and way of life.

Some regulators now regard operational risk as the most important risk facing banks. To quote Thomas J. Curry, head of the Office of the Comptroller of the Currency (OCC) in the United States, in 2012: "Given the complexity of today's banking markets and the sophistication of the technology that underpins it, it is no surprise that the OCC deems operational risk to be high and increasing. Indeed, it is currently at the top of the list of safety and soundness issues for the institutions we supervise." He goes on to argue that operational risk is more important than credit risk.¹ Most banks have always had some framework in place for managing operational risk. However, the prospect of new capital requirements led them to greatly increase the resources they devote to measuring and monitoring operational risk.

It is much more difficult to quantify operational risk than credit or market risk. Operational risk is also more difficult to manage. Financial institutions make a conscious decision to take a certain amount of credit and market risk, and there are many traded instruments that can be used to reduce these risks. Operational risk, by contrast, is a necessary part of doing business. An important part of operational risk management is identifying the types of risk that are being taken and should be insured against. There is always a danger that a huge loss will be incurred from taking an operational risk that ex ante was not even recognized as a risk.

It might be thought that a loss such as that at Société Générale (SocGen, see Business Snapshot 5.5) was a result of market risk because it was movements in market variables that led to it. However, it should be classified as an operational risk loss because it involved fraud. (Jérôme Kerviel created fictitious trades to hide the big bets he was making.) Suppose there was no fraud. If it was part of the bank's policy to let traders take huge risks, then the loss would be classified as a market risk loss. But, if this was not part of the bank's policy and there was a breakdown in its controls, it should be classified as an operational risk loss. The SocGen example illustrates that operational risk losses are often contingent on market movements. If the market had moved in Kerviel's favor, there would have been no loss. The fraud and the breakdown in SocGen's control systems might then never have come to light.

There are some parallels between the operational risk losses of banks and the losses of insurance companies. Insurance companies face a small probability of a large loss from a hurricane, earthquake, or other natural disaster. Similarly, banks face a small probability of a large operational risk loss. But there is one important difference. When insurance

¹Speech to the Exchequer Club, May 16, 2012.

companies lose a large amount of money because of a natural disaster, all companies in the industry tend to be affected and often premiums rise the next year to cover losses. Operational risk losses tend to affect only one bank. Because it operates in a competitive environment, the bank does not have the luxury of increasing prices for the services it offers during the following year.

20.1 Defining Operational Risk

There are many different ways in which operational risk can be defined. It is tempting to consider operational risk as a residual risk and define it as any risk faced by a financial institution that is not market risk or credit risk. To produce an estimate of operational risk, we could then look at the financial institution's financial statements and remove from the income statement (a) the impact of credit losses and (b) the profits or losses from market risk exposure. The variation in the resulting income would then be attributed to operational risk.

Most people agree that this definition of operational risk is too broad. It includes the risks associated with entering new markets, developing new products, economic factors, and so on. Another possible definition is that operational risk, as its name implies, is the risk arising from operations. This includes the risk of mistakes in processing transactions, making payments, and so on. This definition of risk is too narrow. It does not include major risks such as the risk of a rogue trader such as Jérôme Kerviel.

We can distinguish between internal risks and external risks. Internal risks are those over which the company has control. The company chooses whom it employs, what computer systems it develops, what controls are in place, and so on. Some people define operational risks as all internal risks. Operational risk then includes more than just the risk arising from operations. It includes risks arising from inadequate controls such as the rogue trader risk and the risks of other sorts of employee fraud.

Bank regulators favor including more than just internal risks in their definition of operational risk. They include the impact of external events such as natural disasters (for example, a fire or an earthquake that affects the bank's operations), political and regulatory risk (for example, being prevented from operating in a foreign country by that country's government), security breaches, and so on. All of this is reflected in the following definition of operational risk produced by the Basel Committee on Banking Supervision in 2001:

The risk of loss resulting from inadequate or failed internal processes, people, and systems or from external events.

Note that this definition includes legal risk but does not include reputation risk and the risk resulting from strategic decisions.

Operational risks result in increases in the bank's costs or decreases in its revenue. Some operational risks interact with credit and market risk. For example, when mistakes are made in a loan's documentation, it is usually the case that losses result if and only if

the counterparty defaults. When a trader exceeds limits and misreports positions, losses result if and only if the market moves against the trader.

20.2 Types of Operational Risk

The main categories of operational risk (as defined by regulators when they initially chose to require capital for operational risk) are:

1. *Internal fraud*: Acts of a type intended to defraud, misappropriate property, or circumvent regulations, the law, or company policy (excluding those concerned with diversity or discrimination) involving at least one internal party. Examples include intentional misreporting of positions, employee theft, insider trading on an employee's own account, and rogue trading.
2. *External fraud*: Acts by a third party of a type intended to defraud, misappropriate property, or circumvent the law. Examples include robbery, forgery, check kiting, and damage from computer hacking.
3. *Employment practices and workplace safety*: Acts inconsistent with employment, health, or safety laws or agreements, or that result in payment of personal injury claims or claims relating to diversity or discrimination issues. Examples include worker compensation claims, violation of employee health and safety rules, organized labor activities, discrimination claims, and general liability (for example, a customer slipping and falling at a branch office).
4. *Clients, products, and business practices*: Unintentional or negligent failure to meet a professional obligation to clients and the use of inappropriate products or business practices. Examples are fiduciary breaches, misuse of confidential customer information, improper trading activities on the bank's account, money laundering, and the sale of unauthorized products.
5. *Damage to physical assets*: Loss or damage to physical assets from natural disasters or other events. Examples include terrorism, vandalism, earthquakes, fires, and floods.
6. *Business disruption and system failures*: Disruption of business or system failures. Examples include hardware and software failures, telecommunication problems, and utility outages.
7. *Execution, delivery, and process management*: Failed transaction processing or process management, and disputes with trade counterparties and vendors. Examples include data entry errors, collateral management failures, incomplete legal documentation, unapproved access given to clients' accounts, non-client counterparty misperformance, and vendor disputes.

20.3 Loss Severity and Loss Frequency

Two distributions can be important in estimating the potential size of different categories of operational risk. One is the *loss frequency distribution* and the other is the *loss severity*

distribution. The loss frequency distribution is the distribution of the number of losses observed during one year. The loss severity distribution is the distribution of the size of a loss, given that a loss occurs.

For loss frequency, the natural probability distribution to use is a Poisson distribution. This distribution assumes that losses happen randomly through time so that in any short period of time Δt there is a probability $\lambda \Delta t$ of a loss occurring. The probability of n losses in T years is

$$e^{-\lambda T} \frac{(\lambda T)^n}{n!}$$

The parameter λ can be estimated as the average number of losses per year. For example, if during a 10-year period there were a total of 12 losses, λ would be estimated as 1.2 per year. A Poisson distribution has the property that the mean frequency of losses equals the variance of the frequency of losses.²

For the loss-severity probability distribution, a lognormal probability distribution is often used. The parameters of this probability distribution are the mean and standard deviation of the logarithm of the loss.

The loss-frequency distribution can be combined with the loss severity distribution for each risk type/business line combination to determine a loss distribution. Monte Carlo simulation can be used for this purpose.³ (The simplest assumption is that the two distributions are independent.) On each simulation trial, we proceed as follows:

1. We sample from the frequency distribution to determine the number of loss events ($= n$) in one year.
2. We sample n times from the loss severity distribution to determine the loss experienced for each loss event ($L_1, L_2, \dots, L_n$).
3. We determine the total loss experienced ($= L_1 + L_2 + \dots + L_n$).

When many simulation trials are used, we obtain a total loss distribution for losses of the type being considered. The 99.9 percentile of the distribution can then be determined.

Figure 20.1 illustrates the procedure. In this example, the expected loss frequency is 3 per year and the loss severity is drawn from a lognormal distribution. The logarithm of each loss (\$ millions) is assumed to have a mean of zero and a standard deviation of 0.4. The Excel worksheet used to produce Figure 20.1 is on the author's website: www-2.rotman.utoronto.ca/hull/riskman.

²If the mean frequency is greater than the variance of the frequency, a binomial distribution may be more appropriate. If the mean frequency is less than the variance, a negative binomial distribution (mixed Poisson distribution) may be more appropriate.

³Combining the loss severity and loss frequency distribution is a very common problem in insurance. Apart from Monte Carlo simulation, two approaches that are used are Panjer's algorithm and fast Fourier transforms. See H. H. Panjer, "Recursive Evaluation of a Family of Compound Distributions," *ASTIN Bulletin* 12 (1981): 22–29.

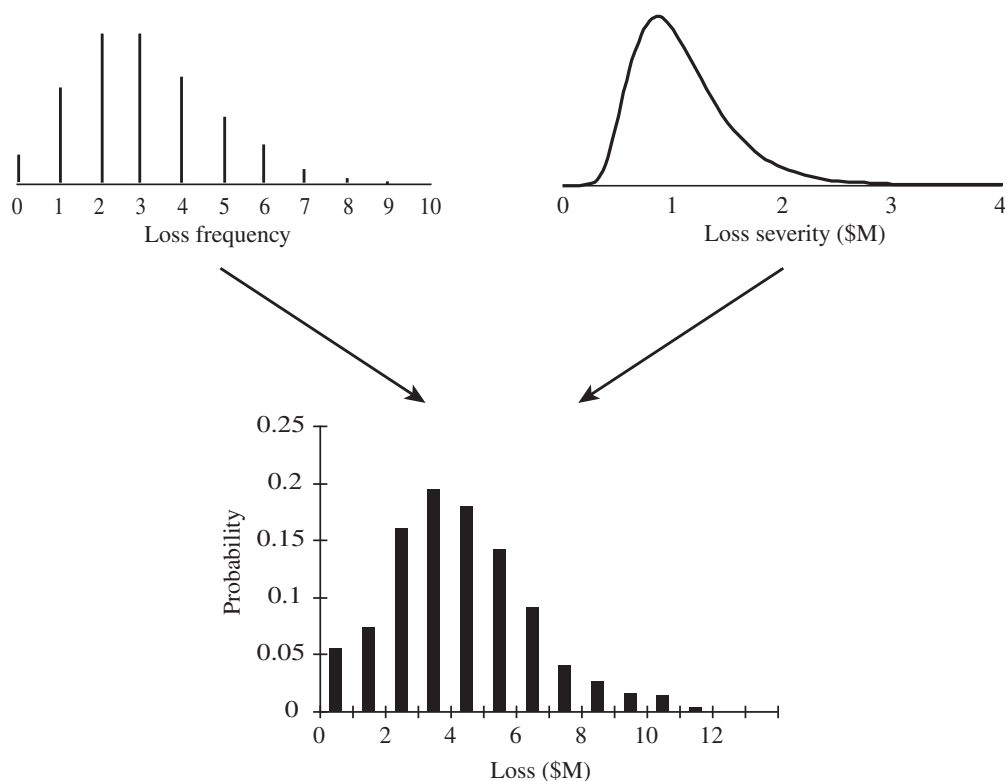


Figure 20.1 Calculation of Loss Distribution from Loss Frequency and Loss Severity

20.3.1 Internal vs. External Data

Unfortunately, there is usually relatively little historical data available within a bank to estimate loss severity and loss frequency distributions for particular types of losses. Many banks have not in the past kept records of operational risk losses. They are doing so now, but it may be some time before a reasonable amount of historical data is available. It is interesting to compare operational risk losses with credit risk losses in this respect. Traditionally, banks have done a much better job at documenting their credit risk losses than their operational risk losses. Also, in the case of credit risks, a bank can rely on a wealth of information published by credit rating agencies to assess probabilities of default and expected losses given default. Similar data on operational risk have not been collected in a systematic way.

Sometimes operational risk losses from other banks (external data) are used. There are two main ways in which this happens. The first involves data consortia, which are companies that facilitate the sharing of data between banks. (The insurance industry has had mechanisms for sharing loss data for many years and banks are now doing this as well.) The second is data vendors, who are in the business of collecting publicly available data in a systematic way. External data increases the amount of data available to a bank for estimating potential losses. It also has the advantage that it can lead to a consideration of

types of losses that have never been incurred by the bank, but which have been incurred by other banks.

Both internal and external historical data must be adjusted for inflation. In addition, a scale adjustment should be made to external data. If a bank with a revenue of \$10 billion reports a loss of \$8 million, how should the loss be scaled for a bank with a revenue of \$5 billion? A natural assumption is that a similar loss for a bank with a revenue of \$5 billion would be \$4 million. But this estimate is probably too small. For example, research by Shih et al. (2000) suggests that the effect of firm size on the size of a loss experienced is non-linear.⁴ Their estimate is

$$\text{Estimated Loss for Bank A} = \text{Observed Loss for Bank B} \times \left(\frac{\text{Bank A Revenue}}{\text{Bank B Revenue}} \right)^\alpha$$

where $\alpha = 0.23$. This means that in our example the bank with a revenue of \$5 billion would experience a loss of $8 \times 0.5^{0.23} = \$6.82$ million. After the appropriate scale adjustment, data obtained through sharing arrangements with other banks can be merged with the bank's own data to obtain a larger sample for determining the loss severity distribution.

The loss data available from data vendors are data taken from public sources such as newspapers and trade journals. Data from vendors cannot be used in the same way as internal data or data obtained through sharing arrangements because they are subject to biases. For example, only large losses are publicly reported, and the larger the loss, the more likely it is to be reported.

Data from vendors are most useful for determining *relative* loss severity. Suppose that a bank has good information on the mean and standard deviation of its loss severity distribution for internal fraud in corporate finance, but not for external fraud in corporate finance or for internal fraud in trading and sales. Suppose that the bank estimates the mean and standard deviation of its loss severity distribution for internal fraud in corporate finance as \$50 million and \$30 million. Suppose further that external data indicate that, for external fraud in corporate finance, the mean severity is twice that for internal fraud in corporate finance and the standard deviation of the severity is 1.5 times as great. In the absence of a better alternative, the bank might assume that its own severity for external fraud in corporate finance has a mean of $2 \times 50 = \$100$ million and a standard deviation of severity equal to $1.5 \times 30 = \$45$ million. Similarly, if the external data indicate that the mean severity for internal fraud in trading and sales is 2.5 times that for internal fraud in corporate finance and the standard deviation is twice as great, the bank might assume that its own severity for internal fraud in trading and sales has a mean of $2.5 \times 50 = \$125$ million and a standard deviation of $2 \times 30 = \$60$ million.

⁴See J. Shih, A. Samad-Khan, and P. Medapa, "Is the Size of an Operational Loss Related to Firm Size?" *Operational Risk Magazine* 2, no. 1 (January 2000). Whether Shih et al.'s results apply to legal risks is debatable. The size of a settlement in a large lawsuit against a bank can be governed by how much the bank can afford.

20.3.2 Scenario Analysis

Scenario analysis is a tool sometimes used for the assessment of operational risk. Some scenarios might come from the bank's own experience, some might be based on the experience of other banks, some might come from the work of consultants, and some might be generated by the risk management group in conjunction with senior management or business unit managers.

An operational risk committee consisting of members of the risk management group and senior management can be asked to estimate loss severity and loss frequency parameters for the scenarios. As explained earlier, a lognormal distribution is often used for loss severity and a Poisson distribution is often used for loss frequency. Data from other banks may be useful for estimating the loss severity parameters. The loss frequency parameters should reflect the controls in place at the bank and the type of business it is doing. They should reflect the views of the members of the operational risk committee. A number of categories of loss frequency can be defined, such as:

1. Scenario happens once every 1,000 years on average ($\lambda = 0.001$)
2. Scenario happens once every 100 years on average ($\lambda = 0.01$)
3. Scenario happens once every 50 years on average ($\lambda = 0.02$)
4. Scenario happens once every 10 years on average ($\lambda = 0.1$)
5. Scenario happens once every 5 years on average ($\lambda = 0.2$)

The committee can be asked to assign each scenario that is developed to one of the categories.

One difference between this scenario analysis and the one in Chapter 16 is that there is no model for determining losses and, if data are not available, the parameters of the loss severity distribution have to be estimated subjectively by the committee. One approach is to ask the committee to estimate an average loss and a high loss that the committee is 99% certain will not be exceeded. A lognormal distribution can then be fitted to the estimates.

Fortunately, the operational risk environment does not usually change as fast as the market and credit risk environment so that the amount of work involved in developing scenarios and keeping them up to date should not be as onerous. Nevertheless, the approach we have described does require a great deal of senior management time. The relevant scenarios for one bank are often similar to those for other banks and, to lessen the burden on the operational risk committee, there is the potential for standard scenarios to be developed by consultants or by bank industry associations. However, the loss frequency estimates should always be specific to a bank and reflect the controls in place in the bank and the type of business it is currently doing.

The advantage of generating scenarios using managerial judgment is that they include losses that the financial institution has never experienced, but could incur. The scenario analysis approach leads to management thinking actively and creatively about potential adverse events. This can have a number of benefits. In some cases, strategies for responding to an event so as to minimize its severity are likely to be developed. In other cases, proposals may be made for reducing the probability of the event occurring at all.

20.4 The Standardized Measurement Approach

In December 2017 the bank regulators finalized rules, referred to as the *standardized measurement approach* (SMA), for calculating capital for operational risk.⁵ The implementation date was January 2023.

Banks must first calculate what is known as the *business indicator* (BI). This is a measure of the bank's size and therefore its potential operational risk exposure. The BI has three components:

1. The interest leases and dividend component (ILDC)
2. The services component (SC)
3. The financial component (FC)

so that

$$BI = ILDC + SC + FC$$

The formulas for the components are:

$$ILDC = \text{Min}[\overline{[\text{Interest Income} - \text{Interest Expense}]}, 0.0225 \times \overline{\text{Interest Earning Assets}}] \\ + \overline{\text{Dividend Income}}$$

$$SC = \text{Max}[\overline{\text{Other Operating Income}}, \overline{\text{Other Operating Expense}}] \\ + \text{Max}[\overline{\text{Fee Income}}, \overline{\text{Fee Expense}}]$$

$$FC = |\overline{\text{Net P\&L Trading Book}}| + |\overline{\text{Net P\&L Banking Book}}|$$

with overbars indicating that averages are taken over the previous three years.

The next stage is to calculate the *business indicator component* (BIC) using the piecewise linear relationship in Table 20.1. The BIC is 12% of BI for small banks. Increases in BI between 1 and 30 billion euros lead to increases in BIC equal to 15% of the BI increase. Increases in BI above 30 billion euros lead to increases in BIC equal to 18% of the BI increase.

The *internal loss multiplier* (ILM) is then calculated as follows:

$$ILM = \text{In} \left[e - 1 + \left(\frac{LC}{BIC} \right)^{0.8} \right]$$

Table 20.1 Calculation of the BIC from the BI

BI (billions of euros)	BIC (billions of euros)
0 to 1	$0.12 \times BI$
1 to 30	$0.12 + 0.15 \times (BI - 1)$
>30	$4.47 + 0.18 \times (BI - 30)$

⁵See www.bis.org/bcbs/publ/d424.pdf.

BUSINESS SNAPSHOT 20.1

The Hammersmith and Fulham Story

Between 1987 and 1989, the London Borough of Hammersmith and Fulham in Great Britain entered into about 600 interest rate swaps and related transactions with a total notional principal of about six billion pounds. The transactions appear to have been entered into for speculative rather than hedging purposes. The two employees of Hammersmith and Fulham who were responsible for the trades had only a sketchy understanding of the risks they were taking and how the products they were trading worked.

By 1989, because of movements in sterling interest rates, Hammersmith and Fulham had lost several hundred million pounds on the swaps. To the banks on the other side of the transactions, the swaps were worth several hundred million pounds. The banks were concerned about credit risk. They had entered into offsetting swaps to hedge their interest rate risks. If Hammersmith and Fulham defaulted, they would still have to honor their obligations on the offsetting swaps and would take a huge loss.

What actually happened was not a default. Hammersmith and Fulham's auditor asked to have the transactions declared void because Hammersmith and Fulham did not have the authority to enter into the transactions. The British courts agreed. The case was appealed and went all the way to the House of Lords, Britain's highest court. The final decision was that Hammersmith and Fulham did not have the authority to enter into the swaps, but that they ought to have the authority to do so in the future for risk management purposes. Needless to say, banks were furious that their contracts were overturned in this way by the courts.

where the *loss component* (LC) is equal to 15 times average annual operational risk losses incurred over the previous 10 years. When high-quality data on losses over 10 years are not available, banks can use losses over five years during a transition period.

For small banks (BI less than one billion euros) the operational risk capital is set equal to BIC. For other banks it is set equal to the product of BIC and ILM. However, national bank supervisors have some discretion here. For example, they can set capital equal to BIC for all banks (while still requiring loss data disclosure). Alternatively, they can use loss data in determining capital for small banks.

To understand the formulas, note that when $LC = BIC$, $ILM=1$, and required capital equals BIC. Regulators expect that this will be true for an "average" bank. Banks experiencing operational risk losses that are higher than average its size will have an ILM

greater than 1, while those experiencing losses less than average will have an ILM less than 1.

20.5 Preventing Operational Risk Losses

A key part of an operational risk manager's job is to be proactive and try to prevent losses before they occur. We now consider some of the ways this can be done.

One approach is to monitor what is happening at other banks and try and learn from their mistakes. When a \$700 million rogue trader loss happened at a Baltimore subsidiary of Allied Irish Bank in 2002, risk managers throughout the world studied the situation carefully and asked: "Could this happen to us?" Business Snapshot 20.1 describes a situation concerning a British local authority in the late 1980s. It immediately led to all banks instituting procedures for checking that counterparties had the authority to enter into derivatives transactions.

20.5.1 Causal Relationships

Operational risk managers should try to establish causal relations between decisions taken and operational risk losses. Does increasing the average educational qualifications of employees reduce losses arising from mistakes in the way transactions are processed? Will a new computer system reduce the probabilities of losses from system failures? Are operational risk losses correlated with the employee turnover rate? Can the risk of a rogue trader be reduced by the way responsibilities are divided between different individuals and by the way traders are motivated?

One approach to establishing causal relationships is statistical. If we look at 12 different locations where a bank operates and find a high negative correlation between the education of back-office employees and the cost of mistakes in processing transactions, it might well make sense to do a cost-benefit analysis of changing the educational requirements for a back-office job in some of the locations. In some cases, a detailed analysis of the cause of losses may provide insights. For example, if 40% of computer failures can be attributed to the fact that the current hardware is several years old and less reliable than newer versions, a cost-benefit analysis of upgrading is likely to be useful.

20.5.2 RCSA and KRIs

Risk control and self-assessment (RCSA) is an important way in which banks try to achieve a better understanding of their operational risk exposures. It involves asking the managers of business units to identify their operational risks. Sometimes questionnaires and scorecards designed by senior management or consultants are used.

A by-product of any program to measure and understand operational risk is likely to be the development of key risk indicators (KRIs).⁶ Risk indicators are key tools in the

⁶These are sometimes referred to as Business Environment and Internal Control Factors (BEICFs).

management of operational risk. The most important indicators are prospective. They provide an early warning system to track the level of operational risk in the organization. Examples of key risk indicators that could be appropriate in particular situations are

1. Staff turnover
2. Number of failed transactions
3. Number of positions filled by temps
4. Ratio of supervisors to staff
5. Number of open positions
6. Percentage of staff who did not take 10 days consecutive leave in the last 12 months

The hope is that key risk indicators can identify potential problems and allow remedial action to be taken before losses are incurred. It is important for a bank to quantify operational risks, but it is even more important to take actions that control and manage those risks.

20.5.3 *E-Mails and Phone Calls*

One way in which operational risk costs can be mitigated is by educating employees to be very careful about what they write in e-mails and, when they work on the trading floor, what they say in phone calls. Lawsuits or regulatory investigations contending that a financial institution has behaved inappropriately or illegally are a major source of operational risk. One of the first things that happens when a case is filed is “discovery,” where the financial institution is required to provide all relevant internal communications. These have often proved to be very embarrassing and have made it difficult for financial institutions to defend themselves. (Fabrice Tourre’s e-mail in Business Snapshot 23.3 is an example of an embarrassing e-mail.) Before sending an e-mail or making a recorded phone call, an employee should consider the question “How could it hurt my employer if this became public knowledge?”

20.6 Allocation of Operational Risk Capital

Operational risk capital should be allocated to business units in a way that encourages them to improve their operational risk management. If a business unit can show that it has taken steps to reduce the frequency or severity of a particular risk, it should be allocated less capital. This will have the effect of improving the business unit’s return on capital (and possibly lead to the business unit manager receiving an increased bonus).

Note that it is not always optimal for a manager to reduce a particular operational risk. Sometimes the costs of reducing the risk outweigh the benefits of reduced capital so that return on allocated capital decreases. A business unit should be encouraged to make appropriate calculations and determine the amount of operational risk that maximizes return on capital.

The overall result of operational risk assessment and operational risk capital allocation should be that business units become more sensitive to the need for managing operational risk. Hopefully, operational risk management will be seen to be an important part of every manager's job. A key ingredient for the success of any operational risk program is the support of senior management. The Basel Committee on Banking Supervision is very much aware of this. It recommends that a bank's board of directors be involved in the approval of a risk management program and that it review the program on a regular basis.

20.7 Use of Power Law

In Section 8.4, we introduced the power law. This states that, for a wide range of variables

$$\text{Prob}(\nu > x) = Kx^{-\alpha}$$

where ν is the value of the variable, x is a relatively large value of ν , and K and α are constants. We covered the theoretical underpinnings of the power law and maximum likelihood estimation procedures when we looked at extreme value theory in Section 12.5.

De Fontnouvelle et al. (2003), using data on losses from vendors, found that the power law holds well for the large losses experienced by banks.⁷ This makes the calculation of VaR with high degrees of confidence, such as 99.9%, easier. Loss data (internal or external) and scenario analysis are employed to estimate the power law parameters using the maximum likelihood approach in Chapter 12. The 99.9 percentile of the loss distribution can then be estimated using equation (12.9).

When loss distributions are aggregated, the distribution with the heaviest tails tends to dominate. This means that the loss with the lowest α defines the extreme tails of the total loss distribution.⁸ Therefore, if all we are interested in is calculating the extreme tail of the total operational risk loss distribution, it may only be necessary to consider one or two business line/risk type combinations.

20.8 Insurance

An important decision for operational risk managers is the extent to which operational risks should be insured against. Insurance policies are available on many different kinds of risk ranging from fire losses to rogue trader losses. In Section 3.7 we discussed the moral hazard and adverse selection risks faced by insurance companies. We now review these again in the context of operational risk.

⁷See P. de Fontnouvelle, V. DeJesus-Rueff, J. Jordan, and E. Rosengren, "Capital and Risk: New Evidence on Implications of Large Operational Risk Losses," *Journal of Money, Credit and Banking* 38, no. 7 (October 2006): 1819–1846.

⁸The parameter ξ in extreme value theory (see Chapter 12) equals $1/\alpha$, so it is the loss distribution with the largest ξ that defines the extreme tails.

BUSINESS SNAPSHOT 20.2

Rogue Trader Insurance

A rogue trader insurance policy presents particularly tricky moral hazard problems. An unscrupulous bank could enter into an insurance contract to protect itself against losses from rogue trader risk and then choose to be lax in its implementation of trading limits. If a trader exceeds the trading limit and makes a large profit, the bank is better off than it would be otherwise. If a large loss results, a claim can be made under the rogue trader insurance policy. Deductibles, coinsurance provisions, and policy limits may mean that the amount recovered is less than the loss incurred by the trader. However, potential net losses to the bank are likely to be far less than potential profits, making the lax trading limits strategy a good bet for the bank.

Given this problem, it is perhaps surprising that some insurance companies do offer rogue trader insurance policies. These companies tend to specify carefully how trading limits are implemented within the bank. They may require that the existence of the insurance policy not be revealed to anyone on the trading floor. They are also likely to want to retain the right to investigate the circumstances underlying any loss.

From the bank's point of view, the lax trading limits strategy we have outlined may be very short-sighted. The bank might well find that the costs of all types of insurance rise significantly as a result of a rogue trader claim. Also, a large rogue trader loss (even if insured) would cause its reputation to suffer.

20.8.1 Moral Hazard

One of the risks facing a company that insures a bank against operational risk losses is moral hazard. This is the risk that the existence of the insurance contract will cause the bank to behave differently than it otherwise would. This changed behavior increases the risks to the insurance company. Consider, for example, a bank that insures itself against robberies. As a result of the insurance policy, it may be tempted to be lax in its implementation of security measures—making a robbery more likely than it would otherwise have been.

Insurance companies have traditionally dealt with moral hazard in a number of ways. Typically there is a *deductible* in any insurance policy. This means that the bank is responsible for bearing the first part of any loss. Sometimes there is a *coinsurance provision* in a policy. In this case, the insurance company pays a predetermined percentage (less than 100%) of losses in excess of the deductible. In addition, there is nearly always a *policy limit*. This is a limit on the total liability of the insurer. Consider again a bank that has

insured itself against robberies. The existence of deductibles, coinsurance provisions, and policy limits is likely to provide an incentive for a bank not to relax security measures in its branches. The moral hazard problem in rogue trader insurance is discussed in Business Snapshot 20.2.

20.8.2 Adverse Selection

The other major problem facing insurance companies is adverse selection. This is where an insurance company cannot distinguish between good and bad risks. It offers the same price to everyone and inadvertently attracts more of the bad risks. For example, banks without good internal controls are more likely to enter into rogue trader insurance contracts; banks without good internal controls are more likely to buy insurance policies to protect themselves against external fraud.

To overcome the adverse selection problem, an insurance company must try to understand the controls that exist within banks and the losses that have been experienced. As a result of its initial assessment of risks, it may not charge the same premium for the same contract to all banks. As time goes by, it gains more information about the bank's operational risk losses and may increase or reduce the premium charged. This is much the same as the approach adopted by insurance companies when they sell automobile insurance to a driver. At the outset, the insurance company obtains as much information on the driver as possible. As time goes by, it collects more information on the driver's risk (number of accidents, number of speeding tickets, etc.) and modifies the premium charged accordingly.

20.9 Sarbanes–Oxley

Largely as a result of the Enron bankruptcy, the Sarbanes–Oxley Act was passed in the United States in 2002. This provides another dimension to operational risk management for both financial and nonfinancial institutions in the United States. The act requires boards of directors to become much more involved with day-to-day operations. They must monitor internal controls to ensure risks are being assessed and handled well.

The act specifies rules concerning the composition of the board of directors of public companies and lists the responsibilities of the board. It gives the SEC the power to censure the board or give it additional responsibilities. A company's auditors are not allowed to carry out any significant non-auditing services for the company.⁹ Audit partners must be rotated. The audit committee of the board must be made aware of alternative accounting treatments. The CEO and CFO must prepare a statement to accompany the audit

⁹Enron's auditor, Arthur Andersen, provided a wide range of services in addition to auditing. The company did not survive the litigation that followed the downfall of Enron.

report to the effect that the financial statements are accurate. The CEO and CFO are required to return bonuses in the event that financial statements are restated. Other rules concern insider trading, disclosure, personal loans to executives, reporting of transactions by directors, and the monitoring of internal controls by directors.

Summary

In 1999, bank supervisors indicated their intention to charge capital for operational risk. This has led banks to carefully consider how they should measure and manage operational risk.

Part of the job of operational risk managers is to try to increase awareness among employees of operational risk so that losses are prevented. They should also try to understand what determines operational risk losses and develop key risk indicators to track the level of operational risk in different parts of the organization.

Once operational risk capital has been estimated, it is important to develop procedures for allocating it to business units. This should be done in a way that encourages business units to reduce operational risk when this can be done without incurring excessive costs.

The power law introduced in Chapter 8 seems to apply to operational risk losses. This makes it possible to use extreme value theory (see Chapter 12) to estimate the tails of a loss distribution from empirical data. When several loss distributions are aggregated, it is the loss distribution with the heaviest tail that dominates. In principle, this makes the calculation of VaR for total operational risk easier.

Many operational risks can be insured against. However, most policies include deductibles, coinsurance provisions, and policy limits. As a result, a bank is always left bearing part of any risk itself. Moreover, the way insurance premiums change as time passes is likely to depend on the claims made and other indicators that the insurance company has of how well operational risks are being managed.

The whole process of measuring, managing, and allocating operational risk is still in its infancy. As time goes by and data are accumulated, improvements to our procedures for managing operational risk are likely to emerge. In many ways, the key benefit of an operational risk management program lies in making managers more aware of the importance of operational risk and perhaps leading to them thinking about it differently.

Further Reading

Bank for International Settlements. "Basel III: Finalising Post-Crisis Reforms," December 2017. www.bis.org/bcbs/publ/d424.htm.

Bank for International Settlements. "Revisions to the Principles for the Sound Management of Operational Risk." March 2021. www.bis.org/bcbs/publ/d515.pdf.

Bank for International Settlements. "Principles for Operational Resilience." March 2021. www.bis.org/bcbs/publ/d516.htm.

- Baud, N., A. Frachot, and T. Roncalli. "Internal Data, External Data and Consortium Data for Operational Risk Management: How to Pool Data Properly." Working Paper, Groupe de Recherche Operationelle, Credit Lyonnais, 2002.
- Brunel, V. "Operational Risk Modelled Analytically." *Risk* 27, no. 7 (July 2014): 55–59.
- Chorafas, D. N. *Operational Risk Control with Basel II: Basic Principles and Capital Requirements*. Oxford, UK: Elsevier, 2003.
- Davis, E., ed. *The Advanced Measurement Approach to Operational Risk*. London: Risk Books, 2006.
- De Fontnouvelle, P., V. DeJesus-Rueff, J. Jordan, and E. Rosengren. "Capital and Risk: New Evidence on Implications of Large Operational Risk Losses." *Journal of Money, Credit, and Banking* 38, no. 7 (October 2006): 1819–1846.
- Dutta, K., and D. Babbel. "Scenario Analysis in the Measurement of Operational Risk Capital: A Change of Measure Approach." *Journal of Risk and Insurance* 81, no. 2 (2014): 303–334.
- Girling, P. X. *Operational Risk Management: A Complete Guide to a Successful Operational Risk Framework*. Hoboken, NJ: John Wiley & Sons, 2013.
- Lambrigger, D. D., P. V. Shevchenko, and M. V. Wüthrich. "The Quantification of Operational Risk Using Internal Data, Relevant External Data, and Expert Opinion." *Journal of Operational Risk* 2, no. 3 (Fall 2007): 3–28.
- McCormack, P., A. Sheen, and P. Umande. "Managing Operational Risk: Moving Towards the Advanced Measurement Approach." *Journal of Risk Management in Financial Institutions* 7, no. 3 (Summer 2014): 239–256.

Practice Questions and Problems (Answers at End of Book)

- 20.1 What risks are included by regulators in their definition of operational risks? What risks are not included?
- 20.2 Suppose that external data show that a loss of \$100 million occurred at a bank with annual revenues of \$1 billion. Your bank has annual revenues of \$3 billion. What is the implication of the external data for losses that could occur at your bank? (Use Shih's result.)
- 20.3 Suppose that there is a 90% probability that operational risk losses of a certain type will not exceed \$20 million. The power law parameter, α , is 0.8. What is the probability of losses exceeding (a) \$40 million, (b) \$80 million, and (c) \$200 million?
- 20.4 Discuss how moral hazard and adverse selection are handled in car insurance.
- 20.5 Give two ways Sarbanes–Oxley affects the CEOs of public companies.
- 20.6 When is a trading loss classified as a market risk and when is it classified as an operational risk?
- 20.7 Discuss whether there is (a) moral hazard and (b) adverse selection in life insurance contracts.
- 20.8 What are external loss data? How are they obtained? How are they used in determining operational risk loss distributions for a bank?
- 20.9 What distributions are commonly used for loss frequency and loss severity?
- 20.10 Give two examples of key risk indicators that might be monitored by a central operational risk management group within a bank.

- 20.11 The worksheet used to produce Figure 20.1 is on the author's website. What is the mean and standard deviation of the loss distribution? Modify the inputs to the simulation to test the effect of changing the loss frequency from three to four.
- 20.12 What are the key inputs to determining regulatory capital under the Standardized Measurement Approach?
- 20.13 Consider two banks, each with a BIC of 200 million euros. Bank A has suffered 10 operational risk losses of 20 million euros in the past 10 years. Bank B has suffered one 200 million euro loss in the past 10 years. What is the operational risk regulatory capital for each bank under SMA?

Further Questions

- 20.14 Suppose that there is a 1% probability that operational risk losses of a certain type exceed \$10 million. Use the power law to estimate the 99.97% worst-case operational risk loss when the α parameter equals (a) 0.25, (b) 0.5, (c) 0.9, and (d) 1.0.
- 20.15 Consider the following two events: (a) A bank loses \$1 billion from an unexpected lawsuit relating to its transactions with a counterparty and (b) an insurance company loses \$1 billion because of an unexpected hurricane in Texas. Suppose that you have the same investment in shares issued by both the bank and the insurance company. Which loss are you more concerned about? Why?
- 20.16 The worksheet used to produce Figure 20.1 is on the author's website. How does the loss distribution change when the loss severity has a beta distribution with upper bound of five, lower bound of zero, and the other parameters both one?
- 20.17 A bank has a BI of 5.5 billion euros. It has had eight operational risk losses in the past 10 years. The amounts of the losses in millions of euros are: 3, 7, 15, 65, 85, 150, 250, and 300. What is the bank's operational risk regulatory capital under SMA?

Chapter 21

Liquidity Risk

The Global Financial Crisis that started in the middle of 2007 has emphasized the importance of liquidity risk for both financial institutions and their regulators. Some financial institutions that relied on wholesale deposits for their funding experienced problems as investors lost confidence in financial institutions. Moreover, financial institutions found that some instruments for which there had previously been a liquid market could only be sold at fire-sale prices during the crisis. A important role of the Treasury function within a financial institution is to manage liquidity risk.

It is important to distinguish solvency from liquidity. Solvency refers to a company having more assets than liabilities, so that the value of its equity is positive. Liquidity refers to the ability of a company to make cash payments as they become due. Financial institutions that are solvent can—and sometimes do—fail because of liquidity problems. Consider a bank whose assets are mostly illiquid mortgages. Suppose the assets are financed 90% with deposits and 10% with equity. The bank is comfortably solvent. But it could fail if there is a run on deposits with 25% of depositors suddenly deciding to withdraw their funds. In this chapter we will examine how Northern Rock, a British bank specializing in mortgage lending, failed largely because of liquidity problems of this type.

It is clearly important for financial institutions to manage liquidity carefully. Liquidity needs are uncertain. A financial institution should assess a worst-case liquidity scenario and make sure that it can survive that scenario by either converting assets into cash or raising cash in some other way. As explained in later chapters, regulators have introduced liquidity requirements that are designed to ensure that banks do this.

Liquidity is also an important consideration in trading. A liquid position in an asset is one that can be unwound at short notice. As the market for an asset becomes less liquid, traders are more likely to take losses because they face bigger bid–ask spreads. For an option or other derivative, it is important for there to be a liquid market for the underlying asset so that the trader has no difficulty in doing the daily trades necessary to maintain delta neutrality (see Chapter 15).

This chapter discusses different aspects of liquidity risk. It considers liquidity trading risk and liquidity funding risk. It also looks at what are termed “liquidity black holes.” These are situations where a shock to financial markets causes liquidity to almost completely dry up.

21.1 Liquidity Trading Risk

If a financial institution owns 100, 1,000, 10,000, or even 100,000 shares in IBM, liquidity risk is not a concern. Several million IBM shares trade on the New York Stock Exchange every day. The quoted price of the shares is very close to the price that the financial institution would be able to sell the shares for. However, not all assets are as readily convertible into cash. For example, a \$100 million investment in the bonds of a non-investment-grade U.S. company might be quite difficult to sell at close to the market price in one day. Shares and debt of companies in emerging markets are likely to be even less easy to sell.

The price at which a particular asset can be sold depends on

1. The mid-market price of the asset, or an estimate of its value
2. How much of the asset is to be sold
3. How quickly it is to be sold
4. The economic environment

When there is a market maker who quotes a bid and ask price for a financial asset, the financial institution can sell the asset at the bid and buy at the ask. (The ask price is sometimes referred to as the offer price.) However, it is usually stated (or understood) that a particular quote is good for trades up to a certain size. Above that size, the market maker is likely to increase the bid–ask spread. This is because the market maker knows that, as the size of a trade increases, the difficulty of hedging the exposure created by the trade also increases.

When there is no market maker for a financial instrument, there is still an implicit bid–ask spread. If a financial institution approaches another financial institution (or an interdealer broker) to do a trade, the price depends on which side of the trade it wants to take. The bid–ask spread for an asset can vary from 0.01% of the asset’s mid-market price to as much as 5%, or even 10%, of its mid-market price.

The general nature of the relationship between bid quotes, ask quotes, and trade size is indicated in Figure 21.1. The bid price tends to decrease and the ask price tends to

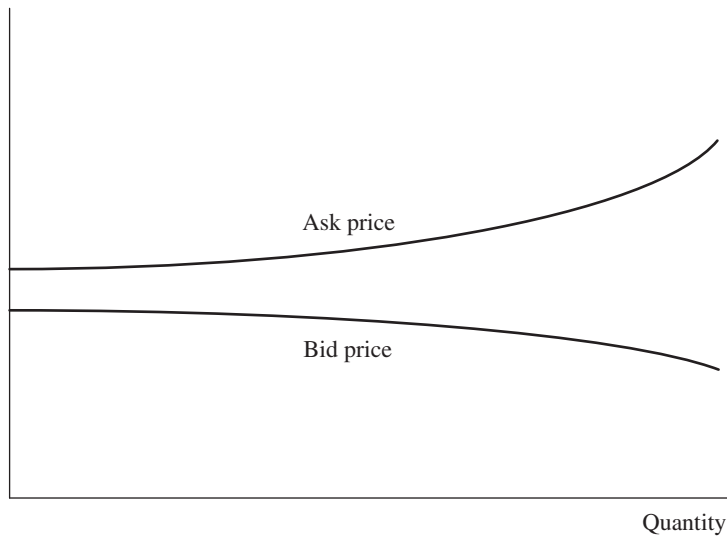


Figure 21.1 Bid and Ask Prices as a Function of Quantity Transacted

increase with the size of a trade. For an instrument where there is a market maker, the bids and asks are the same up to the market maker's size limit and then start to diverge.

Figure 21.1 describes the market for large deals between sophisticated financial institutions. It is interesting to note that bid–ask spreads in the retail market sometimes show the opposite pattern to that in Figure 21.1. Consider, for example, an individual who approaches a branch of a bank wanting to do a foreign exchange transaction or invest money for 90 days. As the size of the transaction increases, the individual is likely to get a better quote.

The price that can be realized for an asset often depends on how quickly it is to be liquidated and the economic environment. Suppose you want to sell your house. Sometimes the real estate market is referred to as a “seller's market.” Almost as soon as you put the house on the market, you can expect to get several offers to purchase and the house will be sold within a week. In other markets, it may take six months or more to sell the house. In the latter case, if you need to sell the house immediately, you will have to reduce the asking price well below the estimated market value.

Financial assets are similar to real assets as far as this is concerned. Sometimes liquidity is tight (e.g., after the Russian default of 1998 and during the Global Financial Crisis). Liquidating even a relatively small position can then be time-consuming and is sometimes impossible. On other occasions, there is plenty of liquidity in the market and relatively large positions can be unwound without difficulty.

Liquidating a large position can be affected by what is termed *predatory trading*. This occurs when a market participant, say Company X, has a large position and other market participants guess that it will have to be unwound in the near future. The other market participants attempt to profit by doing similar trades to those they expect from

Company X. For example, if it is expected that Company X will have to sell a large position in a particular stock, they short the stock in anticipation of a price decline. This makes it more difficult than it would otherwise be for Company X to exit from its position at competitive prices. To avoid predatory trading, financial institutions emphasize to employees the importance of keeping their positions and their future trading plans confidential. Predatory trading was an issue for the trader known as the London Whale in 2012 (see Business Snapshot 22.1) and for Metallgesellschaft (see Business Snapshot 21.3). In the case of Long-Term Capital Management (see Business Snapshot 16.1), positions were unwound slowly over a period of time under the supervision of the Federal Reserve to avoid predatory trading.

Another problem in the market for financial assets is that, when one financial institution finds that it needs to unwind a position, it is often the case that many other financial institutions with similar positions need to do the same thing. The liquidity normally present in the market then evaporates. This is the “liquidity black hole” phenomenon that will be discussed later in this chapter.

21.1.1 The Importance of Transparency

One thing that the market has learned from the Global Financial Crisis is that transparency is important for liquidity. If the nature of an asset is uncertain, it is not likely to trade in a liquid market for very long.

As explained in Chapter 7, it had become common practice in the years prior to 2007 to form portfolios of subprime mortgages and other assets and to create financial instruments by securitizing, re-securitizing, and re-re-securitizing the credit risk. Many of the financial instruments were even more complicated than indicated in Chapter 7. Sometimes the ABS CDOs that were created included non-mortgage assets and even tranches from other ABS CDOs. After August 2007, market participants realized that they knew very little about the risks in the instruments they had traded. Moreover, it was very difficult for them to find out very much about this. Belatedly, they realized they had been using credit ratings as a substitute for an understanding of the instruments.

After August 2007, the instruments created from subprime mortgages became illiquid. Financial institutions had no idea how to mark to market investments that they had been scrambling to buy a few months earlier. They realized that they had purchased highly complicated credit derivatives and that they did not have the tools to value them. They lacked both the necessary models and solid information about the assets in the portfolios underlying the derivatives.

Other well-defined credit derivatives, such as credit default swaps, continued to trade actively during the credit crisis. The lesson from all this is that the market can sometimes get carried away trading complex products that are not transparent, but, when it comes to its senses, liquidity for the products soon disappears. When the products do trade again, prices are likely to be low and bid-ask spreads are likely to be high. As mentioned in Chapter 7, in July 2008 Merrill Lynch agreed to sell \$30.6 billion of ABS CDO tranches (previously rated AAA) to Lone Star Funds for 22 cents on the dollar.

21.1.2 Measuring Market Liquidity

One measure of the market liquidity of an asset is its bid–ask spread. This can be measured either as a cash amount or as a proportion of the asset price. The cash bid–ask spread is

$$p = \text{Ask price} - \text{Bid price}$$

The proportional bid–ask spread for an asset is defined as

$$s = \frac{\text{Ask price} - \text{Bid price}}{\text{Mid-market price}}$$

where the mid-market price is halfway between the bid and the ask price. Sometimes it is convenient to work with the cash bid–ask spread, p , and sometimes with the proportional bid–ask spread, s .

The mid-market price can be regarded as the fair price. In liquidating a position in an asset, a financial institution incurs a cost equal to $s\alpha/2$ where α is the (mid-market) value of the position. This reflects the fact that trades are not done at the mid-market price. A buy trade is done at the ask price and a sell trade is done at the bid price.

One measure of the liquidity of a book is how much it would cost to liquidate the book in normal market conditions within a certain time. Suppose that s_i is an estimate of the proportional bid–ask spread in normal market conditions for the i th financial instrument held by a financial institution and α_i is the value of the position in the instrument. Then

$$\text{Cost of liquidation (normal market)} = \sum_{i=1}^n \frac{s_i \alpha_i}{2} \quad (21.1)$$

where n is the number of positions. Note that although diversification reduces market risk, it does not necessarily reduce liquidity trading risk. However, as explained earlier, s_i increases with the size of position i . Holding many small positions rather than a few large positions therefore tends to entail less liquidity risk. Setting limits to the size of any one position can therefore be one way of reducing liquidity trading risk.

Example 21.1

Suppose that a financial institution has bought 10 million shares of one company and 50 million ounces of a commodity. The shares are bid \$89.5, ask \$90.5. The commodity is bid \$15, ask \$15.1. The mid-market value of the position in the shares is $90 \times 10 = \$900$ million. The mid-market value of the position in the commodity is $15.05 \times 50 = \$752.50$ million. The proportional bid–ask spread for the shares is $1/90$ or 0.01111. The proportional bid–ask spread for the commodity is $0.1/15.05$ or 0.006645. The cost of liquidation in a normal market is

$$900 \times 0.01111/2 + 752.5 \times 0.006645/2 = 7.5$$

or \$7.5 million.

Another measure of liquidity is the cost of liquidation in stressed market conditions within a certain time period. Define μ_i and σ_i as the mean and standard deviation of the proportional bid–ask spread for the i th financial instrument held. Then:

$$\text{Cost of liquidation (stressed market)} = \sum_{i=1}^n \frac{(\mu_i + \lambda \sigma_i) \alpha_i}{2} \quad (21.2)$$

The parameter λ gives the required confidence level for the spread. If, for example, we are interested in considering “worst case” spreads that are exceeded only 1% of the time, and if it is assumed that spreads are normally distributed, then $\lambda = 2.326$.

Example 21.2

Suppose that in Example 21.1 the mean and standard deviation for the bid–ask spread for the shares are \$1.0 and \$2.0, respectively. Suppose further that the mean and standard deviation for the bid–ask spread for the commodity are both \$0.1. The mean and standard deviation for the proportional bid–ask spread for the shares are 0.01111 and 0.02222, respectively. The mean and standard deviation for the proportional bid–ask spread for the commodity are both 0.006645. Assuming the spreads are normally distributed, the cost of liquidation that we are 99% confident will not be exceeded is

$$\begin{aligned} &0.5 \times 900 \times (0.01111 + 2.326 \times 0.02222) \\ &+ 0.5 \times 752.5 \times (0.006645 + 2.326 \times 0.006645) = 36.58 \end{aligned}$$

or \$36.58 million. This is almost five times the cost of liquidation in normal market conditions.

In practice, bid–ask spreads are not normally distributed and it may be appropriate to use a value of λ that reflects their empirical distribution. For example, if it is found that the 99 percentile point of the distribution is 3.6 standard deviations above the mean for a particular category of financial instruments, λ can be set equal to 3.6 for those instruments.

Equation (21.2) assumes that spreads in all instruments are perfectly correlated. This may seem overly conservative, but it is not. When liquidity is tight and bid–ask spreads widen, they tend to do so for all instruments. It makes sense for a financial institution to monitor changes in the liquidity of its book by calculating the measures in equations (21.1) and (21.2) on a regular basis. As we have seen, the bid–ask spread depends on how quickly a position is to be liquidated. The measures in equations (21.1) and (21.2) are therefore likely to be decreasing functions of the time period assumed for the liquidation.

21.1.3 Liquidity-Adjusted VaR

Market value at risk, which we discussed in Chapters 11 to 13, is designed to calculate an estimate of a “worst case” change in the mark-to-market valuation of the trading

book. The measures in equations (21.1) and (21.2) are designed to calculate the cost of liquidating a book if market prices do not change. Although VaR and liquidity risk measures deal with different types of risks, some researchers have suggested combining them into a *liquidity-adjusted VaR* measure. One definition of liquidity-adjusted VaR is regular VaR plus the cost of unwinding positions in a normal market. From equation (21.1) this gives

$$\text{Liquidity-Adjusted VaR} = \text{VaR} + \sum_{i=1}^n \frac{s_i \alpha_i}{2} \quad (21.3)$$

Alternatively it can be defined as regular VaR plus the cost of unwinding positions in a stressed market. From equation (21.2) this gives¹

$$\text{Liquidity-Adjusted VaR} = \text{VaR} + \sum_{i=1}^n \frac{(\mu_i + \lambda \sigma_i) \alpha_i}{2}$$

21.1.4 Unwinding a Position Optimally

A trader wishing to unwind a large position in a financial instrument has to decide on the best trading strategy. If the position is unwound quickly, the trader will face large bid–ask spreads, but the potential loss from the mid-market price moving against the trader is small. If the trader chooses to take several days to unwind the position, the bid–ask spread the trader faces each day will be lower, but the potential loss from the mid-market price moving against the trader is larger.

This type of problem is discussed by Almgren and Chriss (2001).² Suppose that the size of a position is V units and that a trader has to decide how to liquidate it over an n -day period. In this case, it is convenient to define the bid–ask spread as a cash amount rather than as a proportion. Define the cash bid–ask spread when the trader trades q units in one day as $p(q)$. Define q_i as the units traded on day i and x_i as the size of the trader's position at the end of day i ($1 \leq i \leq n$). It follows that $x_i = x_{i-1} - q_i$ for $1 \leq i \leq n$ where x_0 is defined as the initial position size, V .

Each trade costs half the bid–ask spread. The total of the costs related to the bid–ask spread is therefore

$$\sum_{i=1}^n q_i \frac{p(q_i)}{2}$$

Suppose that the mid-market price changes are normally distributed with a standard deviation of σ per day and trading takes place at the beginning of a day. The variance of the change in the value of the trader's position on day i is $\sigma^2 x_i^2$. Assuming that price

¹This was suggested in A. Bangia, F. Diebold, T. Schuermann, and J. Stroughair, "Liquidity on the Outside," *Risk* 12 (June): 68–73.

²See R. Almgren and N. Chriss, "Optimal Execution of Portfolio Transactions," *Journal of Risk* 3 (Winter 2001): 5–39.

changes on successive days are independent, the variance of the change in the value of the position applicable to the unwind is therefore

$$\sum_{i=1}^n \sigma^2 x_i^2$$

A trader might reasonably wish to minimize VaR after trading costs have been considered. This corresponds to minimizing something similar to the liquidity-adjusted VaR measure in equation (21.3). The trader's objective is to choose the q_i so that

$$\lambda \sqrt{\sum_{i=1}^n \sigma^2 x_i^2} + \sum_{i=1}^n q_i \frac{p(q_i)}{2}$$

is minimized subject to

$$\sum_{i=1}^n q_i = V$$

with the x_i being calculated from V and the q_i , as indicated above. The parameter λ measures the confidence level in the VaR estimate. For example, when the confidence level is 99%, and daily price changes are assumed to be normally distributed, $\lambda = 2.326$. Once the $p(q)$ function has been estimated, Excel's Solver can be used for the optimization.

Example 21.3

A trader wishes to unwind a position of 100 million units in an asset over five days. Suppose that the bid–ask spread p (measured in cash) as a function of the daily trading volume is

$$p(q) = a + be^{cq}$$

where $a = 0.1$, $b = 0.05$, and $c = 0.03$ and the amount traded, q , is measured in millions of units.

The standard deviation of the price change per day is \$0.1. A spreadsheet for calculating the optimal strategy can be downloaded from the author's website: www2.rotman.utoronto.ca/~hull/riskman. When the confidence level is 95%, the amounts that should be traded on days 1, 2, 3, 4, and 5 are 48.9, 30.0, 14.1, 5.1, and 1.9 million units, respectively. As the VaR confidence level is reduced, the amounts traded per day show less variability. For example, when the confidence level is 90%, they are 45.0, 29.1, 15.6, 7.0, and 3.3 million units, respectively. When the confidence level is 75% they are 36.1, 26.2, 17.7, 11.6, and 8.4 million units, respectively. In the limit when the confidence level is set equal to 50%, so that the trader is interested only in expected costs, not in the standard deviation of costs, 20 million units should be traded each day.

As this example illustrates, when a position is to be closed out over n days, more than $1/n$ of the position should be traded on the first day. This is because the longer any part of the position is held, the greater the risk of adverse market moves.

21.1.5 Other Measures of Market Liquidity

So far we have focused on bid–ask spread as a measure of market liquidity. Many other measures have been proposed. The volume of trading per day (i.e., the number of times the asset trades in a day) is an important measure. When an asset is highly illiquid, the volume of trading in a day is often zero. The price impact of a trade of a certain size is another measure. A measure somewhat similar to this, but more easily calculated, was proposed by Amihud (2002).³ It is the average of

$$\frac{\text{Absolute value of daily return}}{\text{Daily dollar volume}}$$

over all days in the period considered. This measure is widely used by researchers. Amihud shows that an asset's expected return increases as its liquidity decreases. In other words, investors do get compensated for illiquidity.

21.2 Liquidity Funding Risk

We now move on to consider liquidity funding risk. This is the financial institution's ability to meet its cash needs as they arise. As mentioned at the outset of this chapter, liquidity is not the same as solvency. Financial institutions that are solvent (i.e., have positive equity) can, and sometimes do, fail because of liquidity problems. Northern Rock, a British mortgage lender, is a case in point (see Business Snapshot 21.1).

Liquidity funding problems at a financial institution can be caused by:

1. Liquidity stresses in the economy (e.g., a flight to quality such as that seen during the 2007 to 2009 period). Investors are then reluctant to provide funding in situations where there is any credit risk at all.
2. Overly aggressive funding decisions. There is a tendency for all financial institutions to use short-term instruments to fund long-term needs, creating a liquidity mismatch. Financial institutions need to ask themselves: "How much of a mismatch is too much?"
3. A poor financial performance, leading to a lack of confidence. This can result in a loss of deposits and difficulties in rolling over funding.

Often, when a company experiences severe liquidity problems, all three of these have occurred at the same time. The key to managing liquidity risk is predicting cash needs and ensuring that they can be met in adverse scenarios. Some cash needs are predictable. For example, if a bank has issued a bond, it knows when coupons will have to be paid. Others, such as those associated with withdrawals of deposits by retail customers and drawdowns by corporations on lines of credit that the bank has granted, are less predictable. As the financial instruments entered into by financial institutions have

³See Y. Amihud, "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects," *Journal of Financial Markets* 5 (2002): 31–56.

BUSINESS SNAPSHOT 21.1

Northern Rock

Northern Rock, a British bank, was founded in 1997 when the Northern Rock Building Society floated shares on the London Stock Exchange. In 2007, it was one of the top five mortgage lenders in the United Kingdom. It had 76 branches and offered deposit accounts, savings accounts, loans, and house/contents insurance. The bank grew rapidly between 1997 and 2007. Some of its mortgages were securitized through a subsidiary, Granite, that was based in the Channel Islands.

Northern Rock relied on selling short-term debt instruments for much of its funding. Following the subprime crisis of August 2007, the bank found it very difficult to replace maturing instruments because institutional investors became very nervous about lending to banks that were heavily involved in the mortgage business. The bank's assets were sufficient to cover its liabilities so it was not insolvent. To quote from the Financial Services Authority (FSA) in September 2007: "The FSA judges that Northern Rock is solvent, exceeds its regulatory capital requirement, and has a good quality loan book." But Northern Rock's inability to fund itself was a serious problem. It approached the Bank of England for funding on September 12, 2007, and borrowed about £3 billion from the Tripartite Authority (Bank of England, the Financial Services Authority, and HM Treasury) in the following few days.

On September 13, 2007, the BBC business editor Robert Peston broke the news that the bank had requested emergency support from the Bank of England. On Friday, September 14, there was a run on the bank. Thousands of people lined up for hours to withdraw their funds. This was the first run on a British bank in 150 years. Some customers held their funds in an "Internet-only" account, which they were unable to access due to the volume of customers trying to log on. On Monday, September 17, worried savers continued to withdraw their funds. An estimated £2 billion was withdrawn between September 12 and September 17, 2007.

Depositor insurance in the UK guaranteed 100% of the first £2,000 and 90% of the next £33,000. Late on September 17, 2007, the British Chancellor of the Exchequer, Alistair Darling, announced that the British government and the Bank of England would guarantee all deposits held at Northern Rock. As a result of this announcement and subsequent advertisements in major UK newspapers, the lines outside Northern Rock's branches gradually disappeared. Northern Rock's shares, which had fallen from £12 earlier in the year to £2.67, rose 16% on Mr. Darling's announcement.

During the months following September 12, 2007, Northern Rock's emergency borrowing requirement increased. The Bank of England insisted on a penalty rate of interest to discourage other banks from taking excessive risks. Northern Rock raised some funds by selling assets, but by February 2008 the emergency borrowing reached £25 billion. The bank was then nationalized and the management of the bank was changed. It was split into Northern Rock plc and Northern Rock (Asset Management), with the company's bad debt being put in Northern Rock (Asset Management). In November 2011, Northern Rock plc was bought from the British government for £747 million by the Virgin Group, which is headed by the colorful entrepreneur Sir Richard Branson.

The Northern Rock story illustrates just how quickly liquidity problems can lead to a bank spiraling downward. If the bank had been managed a little more conservatively and had paid more attention to ensuring that it had access to funding, it might have survived.

become more complex, cash needs have become more difficult to predict. For example, downgrade triggers (see Section 18.2.3), guarantees provided by a financial institution, and possible defaults by counterparties in derivatives transactions can have an unexpected impact on cash resources.

21.2.1 Sources of Liquidity

The main sources of liquidity for a financial institution are:

1. Holdings of cash and Treasury securities
2. The ability to liquidate trading book positions
3. The ability to borrow money at short notice
4. The ability to offer favorable terms to attract retail and wholesale deposits at short notice
5. The ability to securitize assets (such as loans) at short notice
6. Borrowings from the central bank

We now consider each of these in turn.

Cash and Treasury Securities Cash and Treasury securities are excellent sources of liquidity. Cash is of course always available to meet liquidity needs and Treasury securities issued by countries such as the United States and the United Kingdom can generally be converted into cash at short notice without any problem. However, cash and Treasury securities are relatively expensive sources of liquidity. There is a trade-off between the liquidity of an asset and the return it provides. In order to be profitable, a financial institution needs to invest in assets such as loans to corporations that provide

a higher rate of return than Treasury instruments. There is therefore a limit to the cash and Treasury securities that can reasonably be held.

Liquidating Trading Book Positions Liquidity funding risk is related to liquidity trading risk, considered in Section 21.1, because one way a financial institution can meet its funding requirements is by liquidating part of its trading book. It is therefore important for a financial institution to quantify the liquidity of its trading book so that it knows how easy it would be to use the book to raise cash. The financial institution wants to make sure that it will be able to survive stressed market conditions where there is a general shortage of liquidity. The financial institution's analysis should therefore be based on stressed market conditions, not normal market conditions. This is the reason why the measures discussed in the previous section concerned with bid–ask spreads in stressed markets are important.

Ability to Borrow When markets are unstressed, a creditworthy bank usually has no problem in borrowing money, but in stressed market conditions there is a heightened aversion to risk. This leads to higher interest rates, shorter maturities for loans, and in some cases a refusal to provide funds at all. Financial institutions should monitor the assets that can be pledged as collateral for loans at short notice. A financial institution can (at a cost) mitigate its funding risks somewhat by arranging lines of credit. For example, Countrywide, an originator of mortgages in the United States, had a syndicated loan facility of \$11.5 billion, which it was able to use during the credit crisis of 2007. (This helped keep the company afloat, but it still experienced significant problems and was taken over by Bank of America in January 2008.) As Business Snapshot 21.1 shows, Northern Rock, a similar British mortgage lender, did not fare so well.

Wholesale and Retail Deposits Wholesale deposits are a more volatile source of funding than retail deposits and can disappear quickly in stressed market conditions. Even retail deposits are not as stable as they used to be because it is very easy to compare interest rates offered by different financial institutions and make transfers via the Internet. Unfortunately, liquidity problems tend to be market-wide rather than something that affects one or two financial institutions. When one financial institution wants to increase its retail or wholesale deposit base for liquidity reasons by offering more attractive rates of interest, others usually want to do the same thing and the increased funding is likely to be difficult to achieve.

Securitization As mentioned in Chapter 2, banks have found the “originate-to-distribute” model attractive. Rather than keep illiquid assets such as loans on their balance sheet, they have securitized them. The structures developed for doing this were discussed in Chapter 7. Prior to August 2007, securitization was an important source of liquidity for banks. However, this source of liquidity dried up almost overnight in August 2007 as investors decided that the securitized products were too risky. “Originate-to-distribute” was no longer a viable strategy, and banks had to fund their loans. Not surprisingly, banks became a lot less willing to lend.

Securitization led to other liquidity problems in August 2007. Banks had entered into liquidity backstop arrangements on the asset-backed commercial paper (ABCP) that was used to fund debt instruments, such as mortgages, prior to their securitization. When buyers could not be found, they had to buy the instruments themselves. In some cases, in order to avoid their reputations being adversely impacted, they had to provide financial support to conduits and other off-balance-sheet vehicles that were involved in securitization, even though they were not legally required to do so.

Central Bank Borrowing Central banks (e.g., the Federal Reserve Board in the United States, the Bank of England in the UK, or the European Central Bank) are often referred to as “lenders of last resort.” When commercial banks are experiencing financial difficulties, central banks are prepared to lend money to maintain the health of the financial system. Collateral has to be posted by the borrowers and the central bank typically applies a haircut (i.e., it lends less than 100% of the value of the collateral) and may charge a relatively high rate of interest. In March 2008, after the failure of Bear Stearns (which was taken over by JPMorgan Chase), the Federal Reserve Board extended its borrowing facility to investment banks as well as commercial banks.⁴ Later, it also made the facility available to Fannie Mae and Freddie Mac (which were taken over by the government in September 2008).

Different central banks apply different rules. Following the credit crisis of August 2007, the haircuts used by the European Central Bank (ECB) were lower than those of other central banks. As a result, some British banks preferred to borrow from the European Central Bank (ECB) rather than the Bank of England. (There are even stories of North American banks contemplating the setting up of subsidiaries in Ireland to access the ECB.) By September 2008, the ECB had lent 467 billion euros and announced that it would apply larger haircuts in the future.

Banks try to keep their borrowing from a central bank a secret. There is a danger that the use of central bank borrowings will be interpreted by the market as a sign that the bank is experiencing financial difficulties with the result that other sources of liquidity dry up. As Business Snapshot 21.1 discusses, news that Northern Rock required emergency borrowing led to an immediate run on the bank, exacerbating its liquidity problems.

Hedging Issues Liquidity problems are liable to arise when companies hedge illiquid assets with contracts that are subject to margin requirements. As indicated in Business Snapshot 8.1, gold-mining companies often hedge their risks by entering into agreements with financial institutions to sell gold forward for two or three years. Often the gold-mining company is required to post margin and the amount of the margin required is calculated every day to reflect the value of its forward agreements. If the price

⁴Central banks are concerned about the failure of investment banks because of systemic risk (see Business Snapshot 15.1). Investment banks have derivatives contracts with other investment banks and with commercial banks. There is a danger that, because of the huge amount of trading between banks, a failure by an investment bank will have a ripple effect throughout the financial sector leading to a failure by commercial banks.

BUSINESS SNAPSHOT 21.2

Ashanti Goldfields

Ashanti Goldfields, a West African gold-mining company based in Ghana, experienced problems resulting from its hedging program in 1999. It had sought to protect its shareholders from gold price declines by selling gold forward. On September 26, 1999, 15 European central banks surprised the market with an announcement that they would limit their gold sales over the following five years. The price of gold jumped up over 25%. Ashanti was unable to meet margin calls and this resulted in a major restructuring, which included the sale of a mine, a dilution of the interest of its equity shareholders, and a restructuring of its hedge positions.

of gold rises fast, the forward agreements lose money and result in big margin calls being made by the financial institution on the gold-mining company. The losses on the forward agreements are offset by increases in the value of the gold in the ground—but this is an illiquid asset. As discussed in Business Snapshot 21.2, Ashanti Goldfields was forced to undertake a major restructuring when it could not meet margin calls after a sharp rise in the price of gold.

Another extreme example of a liquidity funding problem caused by hedging is provided by a German company, Metallgesellschaft, that entered into profitable fixed-price oil and gas contracts with its customers (see Business Snapshot 21.3). The lesson from the Ashanti and Metallgesellschaft episodes is not that companies should not use forward and futures contracts for hedging, but rather that they should ensure that they have access to funding to handle the cash flow mismatches that might arise in extreme circumstances.

21.2.2 Reserve Requirements

Reserve requirements require that a proportion of a bank's deposits be kept as cash or on deposit with the central bank. Some countries, such as Canada and the United Kingdom, have no compulsory reserve requirements. The United States eliminated reserve requirements in March 2020. In China, reserve requirements were set at 8% in December 2021.

In addition to ensuring that banks keep a minimum amount of liquidity, reserve requirements affect the money supply. When the reserve requirement is 10%, a \$100 deposit leads to \$90 of lending, which leads to a further \$90 of deposits in the banking system, which leads to further \$81 of lending, and so on. As this process continues, the total money supply (M1) that is created is $90 + 81 + 72.9 + \dots$ or \$900. If the reserve requirement is 20%, a \$100 deposit leads to \$80 of lending, which leads to \$64 of lending, and so on. The total increase in the money supply is $80 + 64 + 51.2 + \dots$ or \$400.

BUSINESS SNAPSHOT 21.3

Metallgesellschaft

In the early 1990s, Metallgesellschaft (MG) sold a huge volume of 5- to 10-year heating oil and gasoline fixed-price supply contracts to its customers at six to eight cents above market prices. It hedged its exposure with long positions in short-dated futures contracts that were rolled forward. As it turned out, the price of oil fell and there were margin calls on the futures positions. MG's trading was made more difficult by the fact that its trades were very large and were anticipated by others.

Considerable short-term cash flow pressures were placed on MG. The members of MG who devised the hedging strategy argued that these short-term cash outflows were offset by positive cash flows that would ultimately be realized on the long-term fixed-price contracts. However, the company's senior management and its bankers became concerned about the huge cash drain. As a result, the company closed out all the hedge positions and agreed with its customers that the fixed-price contracts would be abandoned. The outcome was a loss to MG of \$1.33 billion.

21.2.3 Regulation

As will be explained in Chapter 26, as part of the Basel III, bank regulators have introduced two liquidity risk requirements: the liquidity coverage ratio (LCR) and the net stable funding ratio (NSFR).

The LCR requirement is

$$\frac{\text{High-quality liquid assets}}{\text{Net cash outflows in a 30-day period}} \geq 100\%$$

The 30-day period considered in the calculation of LCR is one of acute stress involving a downgrade of three notches (e.g., from AA+ to A+), a partial loss of deposits, a complete loss of wholesale funding, increased haircuts on secured funding, and drawdowns on lines of credit.

The NSFR requirement is

$$\frac{\text{Amount of stable funding}}{\text{Required amount of stable funding}} \geq 100\%$$

The numerator is calculated by multiplying each category of funding (capital, wholesale deposits, retail deposits, etc.) by an available stable funding (ASF) factor, reflecting their stability (see Table 26.4). The denominator is calculated from the assets and

off-balance-sheet items requiring funding. Each category of these is multiplied by a required stable funding (RSF) factor to reflect the permanence of the funding (see Table 26.5).

Bank regulators have issued the following principles for the management of liquidity risk.⁵

1. A bank is responsible for the sound management of liquidity risk. A bank should establish a robust liquidity risk management framework that ensures it maintains sufficient liquidity, including a cushion of unencumbered, high-quality liquid assets, to withstand a range of stress events, including those involving the loss or impairment of both unsecured and secured funding sources. Supervisors should assess the adequacy of both a bank's liquidity risk management framework and its liquidity position and should take prompt action if a bank is deficient in either area in order to protect depositors and to limit potential damage to the financial system.
2. A bank should clearly articulate a liquidity risk tolerance that is appropriate for its business strategy and its role in the financial system.
3. Senior management should develop a strategy, policies, and practices to manage liquidity risk in accordance with the risk tolerance and to ensure that the bank maintains sufficient liquidity. Senior management should continuously review information on the bank's liquidity developments and report to the board of directors on a regular basis. A bank's board of directors should review and approve the strategy, policies, and practices related to the management of liquidity at least annually and ensure that senior management manages liquidity risk effectively.
4. A bank should incorporate liquidity costs, benefits, and risks in the internal pricing, performance measurement, and new product approval process for all significant business activities (both on- and off-balance-sheet), thereby aligning the risk-taking incentives of individual business lines with the liquidity risk exposures their activities create for the bank as a whole.
5. A bank should have a sound process for identifying, measuring, monitoring, and controlling liquidity risk. This process should include a robust framework for comprehensively projecting cash flows arising from assets, liabilities, and off-balance-sheet items over an appropriate set of time horizons.
6. A bank should actively monitor and control liquidity risk exposures and funding needs within and across legal entities, business lines, and currencies, taking into account legal, regulatory, and operational limitations to the transferability of liquidity.
7. A bank should establish a funding strategy that provides effective diversification in the sources and tenor of funding. It should maintain an ongoing presence in its chosen funding markets and strong relationships with funds providers to promote effective diversification of funding sources. A bank should regularly gauge its

⁵See Bank for International Settlements, "Principles for Sound Liquidity Risk Management and Supervision," September 2008.

capacity to raise funds quickly from each source. It should identify the main factors that affect its ability to raise funds and monitor those factors closely to ensure that estimates of fund-raising capacity remain valid.

8. A bank should actively manage its intraday liquidity positions and risks to meet payment and settlement obligations on a timely basis under both normal and stressed conditions and thus contribute to the smooth functioning of payment and settlement systems.
9. A bank should actively manage its collateral positions, differentiating between encumbered and unencumbered assets. A bank should monitor the legal entity and physical location where collateral is held and how it may be mobilized in a timely manner.
10. A bank should conduct stress tests on a regular basis for a variety of short-term and protracted institution-specific and market-wide stress scenarios (individually and in combination) to identify sources of potential liquidity strain and to ensure that current exposures remain in accordance with a bank's established liquidity risk tolerance. A bank should use stress test outcomes to adjust its liquidity risk management strategies, policies, and positions and to develop effective contingency plans.
11. A bank should have a formal contingency funding plan (CFP) that clearly sets out the strategies for addressing liquidity shortfalls in emergency situations. A CFP should outline policies to manage a range of stress environments, establish clear lines of responsibility, include clear invocation and escalation procedures, and be regularly tested and updated to ensure that it is operationally robust.
12. A bank should maintain a cushion of unencumbered, high-quality liquid assets to be held as insurance against a range of liquidity stress scenarios, including those that involve the loss or impairment of unsecured and typically available secured funding sources. There should be no legal, regulatory, or operational impediment to using these assets to obtain funding.
13. A bank should publicly disclose information on a regular basis that enables market participants to make an informed judgment about the soundness of its liquidity risk management framework and liquidity position.

Recommendations for banks supervisors are:

14. Supervisors should regularly perform a comprehensive assessment of a bank's overall liquidity risk management framework and liquidity position to determine whether they deliver an adequate level of resilience to liquidity stress given the bank's role in the financial system.
15. Supervisors should supplement their regular assessments of a bank's liquidity risk management framework and liquidity position by monitoring a combination of internal reports, prudential reports, and market information.
16. Supervisors should intervene to require effective and timely remedial action by a bank to address deficiencies in its liquidity risk management processes or liquidity position.

17. Supervisors should communicate with other supervisors and public authorities, such as central banks, both within and across national borders, to facilitate effective cooperation regarding the supervision and oversight of liquidity risk management. Communication should occur regularly during normal times, with the nature and frequency of the information sharing increasing as appropriate during times of stress.

21.3 Liquidity Black Holes

It is sometimes argued that technological and other developments have led to an improvement in the liquidity of financial markets. This is questionable. It is true that bid–ask spreads have on average declined. But there has also been an increasing tendency for situations to develop where almost everyone wants to do the same type of trade at the same time. The result has been that what are referred to as “liquidity black holes” occur with increasing frequency.⁶ A liquidity black hole describes a situation where liquidity has dried up in a particular market because everyone wants to sell and no one wants to buy. It is sometimes also referred to as a “crowded exit.”⁷

In a well-functioning market, the market may change its opinion about the price of an asset because of new information. However, the price does not overreact. If a price decrease is too great, traders will quickly move in and buy the asset and a new equilibrium price will be established. A liquidity black hole is created when a price decline causes more market participants to want to sell, driving prices well below where they will eventually settle. During the sell-off, liquidity dries up and the asset can be sold only at a fire-sale price.⁸

21.3.1 Positive and Negative Feedback Traders

Changes in the liquidity of financial markets are driven by the behavior of traders. There are two sorts of traders in the market: negative feedback traders and positive feedback traders.⁹ Negative feedback traders buy when prices fall and sell when prices rise; positive feedback traders sell when prices fall and buy when prices rise.

In liquid markets, negative feedback traders dominate the trading. If the price of an asset gets unreasonably low, traders will move in and buy. This creates demand for the asset that restores the price to a more reasonable level. Similarly, if the price of an asset gets unreasonably high, traders will sell. This creates supply of the asset that also

⁶See A. D. Persaud, ed., *Liquidity Black Holes: Understanding, Quantifying and Managing Financial Liquidity Risk* (London: Risk Books, 1999).

⁷See, for example, J. Clunie, *Predatory Trading and Crowded Exits: New Thinking on Market Volatility* (Petersfield, UK: Harriman House, 2010).

⁸Liquidity black holes tend to be associated with price decreases, but it is in theory also possible for them to occur when there are price increases.

⁹This is a simplification of reality to help understand the dynamics of markets. Some traders follow complicated strategies that cannot be classified as positive feedback or negative feedback.

restores the price to a more reasonable level. The result is that the market is liquid with reasonable prices and a good balance of buyers and sellers.

When positive feedback traders dominate the trading, market prices are liable to be unstable and the market may become one-sided and illiquid. A reduction in the price of an asset causes traders to sell. This results in prices falling further and more selling. An increase in the price of an asset causes traders to buy. This results in the price of the asset increasing further and more buying.

There are a number of reasons why positive feedback trading exists. For example:

1. *Trend trading.* Trend traders attempt to identify trends in an asset price. They buy when the asset price appears to be trending up and sell when it appears to be trending down. A related strategy is *breakout trading*, which involves trading when an asset's price moves outside an established range. For example, if a stock has been trading in the range of \$25 to \$30 for the last six months, traders might be inclined to buy it as soon as it moves above \$30 and sell it as soon as it moves below \$25.
2. *Stop-loss rules.* Traders often have rules to limit their losses. When the price of an asset that is owned falls below a certain level, they automatically sell to limit their losses. These rules are known as "stop-loss" rules and are a source of positive feedback trading that is always present in the market.
3. *Dynamic hedging.* Chapter 15 explains how options traders maintain a delta-neutral position using dynamic hedging. In particular, Tables 15.2 and 15.3 show how a trader would hedge a short position in an option over a 20-week period. Hedging a short option position (call or put) involves buying after a price rise and selling after a price decline. This is positive feedback trading that has the potential to reduce liquidity. (By contrast, dynamically hedging a long position in a call or put option involves selling after a price rise and buying after a price decline. This is negative feedback trading and should not interfere with market liquidity.) Any situation where banks have a large short option position has the potential to destabilize the market and cause illiquidity. As discussed in Business Snapshot 3.1, at one point banks sold a huge volume of options on long-term interest rates to British insurance companies. As the banks hedged their risks, the behavior of long-term interest rates in the UK was dramatically affected.
4. *Creating options synthetically.* Hedging a short position in an option is equivalent to creating a long position in the same option synthetically. It follows that a financial institution can create a long option position synthetically by doing the same sort of trading as it would do if it were hedging a short option position. This leads to positive feedback trading that can cause market instability and illiquidity. The classic example here is the stock market crash of October 1987. In the period leading up to the crash, the stock market had done very well. Increasing numbers of portfolio managers were using commercially available programs to synthetically create put options on their portfolios. These programs told them to sell part of their portfolio immediately after a price decline and buy it back immediately after a price increase. The result, as indicated in Business Snapshot 21.4, was a liquidity black hole where prices plunged on October 19, 1987. In this case, the liquidity black hole was relatively short-lived. Within four months the market recovered to close to its pre-crash level.

BUSINESS SNAPSHOT 21.4

The Crash of 1987

On Monday, October 19, 1987, the Dow Jones Industrial Average dropped by more than 20%. Portfolio insurance played a major role in this crash. In October 1987, portfolios involving over \$60 billion of equity assets were being managed with trading rules that were designed to synthetically create put options on the portfolios. These trading rules involved selling equities (or selling index futures) when the market declined and buying equities (or buying equity futures) when the market rose.

During the period Wednesday, October 14, 1987, to Friday, October 16, 1987, the market declined by about 10%, with much of this decline taking place on Friday afternoon. The portfolio insurance rules should have generated at least \$12 billion of equity or index futures sales as a result of this decline. In fact, portfolio insurers had time to sell only \$4 billion and they approached the following week with huge amounts of selling already dictated by their models. It is estimated that on Monday, October 19, sell programs by three portfolio insurers accounted for almost 10% of the sales on the New York Stock Exchange, and that portfolio insurance sales amounted to 21.3% of all sales in index futures markets. It is likely that the decline in equity prices was exacerbated by investors other than portfolio insurers selling heavily because they anticipated the actions of portfolio insurers.

Because the market declined so fast and the stock exchange systems were overloaded, many portfolio insurers were unable to execute the trades generated by their models and failed to obtain the protection they required. Needless to say, the popularity of portfolio insurance schemes has declined significantly since 1987. One of the lessons from this story is that it is dangerous to follow a particular trading strategy—even a hedging strategy—when many other market participants are doing the same thing. To quote from a report on the crash, “Liquidity sufficient to absorb the limited selling demands of investors became an illusion of liquidity by massive selling, as everyone showed up on the same side of the market at once. Ironically, it was this illusion of liquidity which led certain similarly motivated investors, such as portfolio insurers, to adopt strategies which call for liquidity far in excess of what the market could supply.”

5. *Margins.* A big movement in market variables, particularly for traders who are highly leveraged, may lead to margin calls that cannot be met. This forces traders to close out their positions, which reinforces the underlying move in the market variables. It is likely that volatility increases. This may exacerbate the situation because it leads to exchanges increasing their margin requirements.

6. *Predatory trading.* This was mentioned in Section 21.1. If traders know that an investor must sell large quantities of a certain asset, they know that the price of the asset is likely to decrease. They therefore short the asset. This reinforces the price decline and results in the price falling even further than it would otherwise do. To avoid predatory trading, large positions must usually be unwound slowly.
7. *LTCM.* The failure of the hedge fund Long-Term Capital Management (LTCM), as outlined in Business Snapshot 16.1, provides an example of positive feedback trading. One type of LTCM's trade was "relative value fixed income." LTCM would take a short position in a liquid bond and a long position in a similar illiquid bond, and wait for the prices to move close together. After the Russian default in 1998, the prices of illiquid instruments declined relative to similar liquid instruments. LTCM (and other companies that were following similar strategies to LTCM) were highly leveraged and unable to meet margin calls. They were forced to close out their positions. This involved buying the liquid bonds and selling the illiquid bonds. This reinforced the flight to quality and made the prices of illiquid and liquid bonds diverge even further.

21.3.2 Leveraging and Deleveraging

A phenomenon in the market is leveraging and deleveraging. This is illustrated in Figures 21.2 and 21.3. When banks are awash with liquidity (e.g., because they have developed ways of securitizing assets or because deposit levels are higher than usual), they make credit easily available to businesses, investors, and consumers. Credit spreads decrease. The easy availability of credit increases demand for both financial and nonfinancial assets and the prices of these assets rise. Assets are often pledged as collateral for the loans that are used to finance them. When the prices of the assets rise, the collateral underlying loans (when measured at market prices) is greater and borrowing can increase further. This leads to further asset purchases and a repeat of the cycle. This cycle is referred to as "leveraging" because it leads to more borrowing throughout the economy.

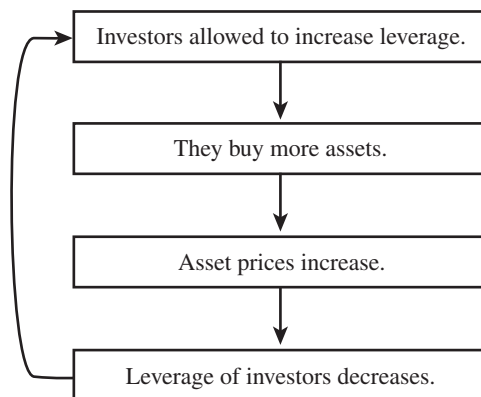


Figure 21.2 Leveraging

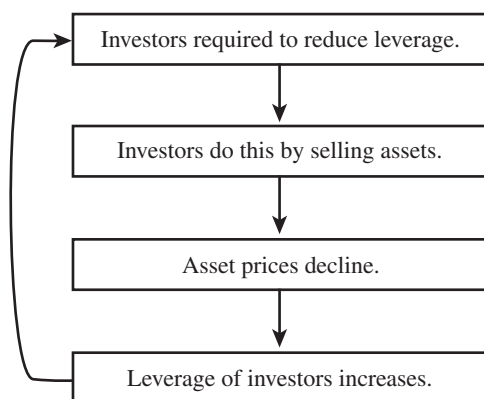


Figure 21.3 Deleveraging

Deleveraging is the opposite process to leveraging. Banks find themselves less liquid for some reason (e.g., because there is less demand for the products of securitization). They become more reluctant to lend money. Credit spreads increase. There is less demand for both nonfinancial and financial assets and their prices decrease. The value of the collateral supporting loans decreases and banks reduce lines of credit. This leads to asset sales being necessary and a further reduction in asset prices.

The period leading up to 2007 was characterized by leveraging for many of the world's economies. Credit spreads declined and it was relatively easy to borrow money for a wide range of different purposes. From the middle of 2007 onward, the situation changed and the deleveraging process shown in Figure 21.3 started. Credit spreads increased, it became much less easy to borrow money, and asset prices decreased.

Hedge funds are particularly affected by the leveraging–deleveraging cycle. Consider a hedge fund that is able to borrow 20 times its equity during the pre-2007 period. Soon after the middle of 2007, the hedge fund might get a call from its prime broker telling it to reduce leverage to, say, five times equity. It can only do this by selling assets. Asset prices decrease as a result of what this hedge fund, and other hedge funds, are doing. The hedge fund's equity declines and further sales are necessary.

21.3.3 Irrational Exuberance

The term “irrational exuberance” was used by Alan Greenspan, Federal Reserve Board chairman, in a speech in December 1996 when, in reference to the stock market, he asked, “How do we know when irrational exuberance has unduly escalated asset values?” (The phrase has been remembered because the speech was followed by declines in stock prices worldwide.) Most liquidity black holes can be traced to irrational exuberance of one sort or another. What happens is that traders working for many different financial institutions become irrationally exuberant about a particular asset class or a particular market variable. The balance sheets of financial institutions then become overextended

through the accumulation of exposure to this asset class or market variable. Often the process is self-reinforcing. When many financial institutions choose to take a particular position, prices increase, making the position look profitable. This creates extra desire on the part of financial institutions to take the position and yet more profits. Risk managers working for the financial institution should (and probably will) complain about the risks being taken, but in many instances senior management are likely to ignore their concerns because high profits are being made. To quote Chuck Prince, ex-CEO of Citigroup, on July 10, 2007: “When the music stops, in terms of liquidity, things will be complicated. But as long as the music is playing, you’ve got to get up and dance. We’re still dancing.”

At some stage the bubble must burst. Many traders then try to get out of their positions at the same time, causing illiquid markets and huge losses. Volatility increases and the risk management procedures used within the financial institution (e.g., the calculation of market VaR from historical data) can cause many financial institutions to try to unwind a wide range of risky positions at the same time. This can lead to further losses and more serious illiquidity problems. There may be failures (or rumors of failures) by some banks. Most banks are likely to experience liquidity funding problems and as a result lending may be curtailed.

The classic example of what has been described above is the Global Financial Crisis. Other examples are the 1987 stock market crash, the 1994 bond market crash, the 1997–1998 Asian monetary crisis, and the 1998 Long-Term Capital Management failure. Irrational exuberance is part of human nature and to some extent is inevitable. As discussed in Chapter 7, it is exacerbated by the way traders are paid. A large part of the compensation comes from a bonus at year-end, which depends on performance during the year. A trader may be savvy enough to know that a market is irrationally exuberant and that there will be a correction. However, if there is a good chance that the correction will be delayed until next year, the trader is motivated to continue building up his or her position to maximize short-term compensation.

21.3.4 The Impact of Regulation

Regulators try to ensure that banks throughout the world are regulated in the same way. However, a uniform regulatory environment comes with costs. All banks tend to respond in the same way to external events. Consider for example market risk. When volatilities and correlations increase, market risk VaR and the capital required for market risks increase. Banks then take steps to reduce their exposures. Since banks often have similar positions to each other, they try to do similar trades. A liquidity black hole can develop.

There is a similar issue as far as credit risk is concerned. During the low point of the economic cycle, default probabilities are relatively high and capital requirements for loans tend to be high. As a result, banks may be less willing to make loans, creating problems for small and medium-sized businesses. During the high point of the business cycle, default

probabilities are relatively low and banks may be too willing to grant credit. (This is similar to the phenomenon described in Figures 21.2 and 21.3.) One way regulators have dealt with this problem is by using an average of the probability of default through the economic or credit cycle, rather than an estimate applicable to one particular point in time.

Should other financial institutions such as life insurance companies and pension funds be regulated in the same way as banks? It is tempting to answer “yes” so that one financial institution is not given an advantage over others. But the answer should be “no.” These financial institutions have longer time horizons than banks. They should not be penalized for investing in illiquid assets. Also, they should not be required to adjust their portfolios when volatilities and correlations increase. These parameters tend to be mean reverting and so they eventually decrease again.

21.3.5 The Importance of Diversity

Models in economics sometimes assume that market participants act independently of each other. We have argued that this is often not the case. It is this lack of independence that causes liquidity black holes. Traders working for financial institutions tend to want to do the same trades at the same time. To solve the problem of liquidity black holes, we need more diversity in financial markets. One way of creating diversity is to recognize that different types of financial institutions have different types of risks and should be regulated differently.

Hedge funds have become important market participants. They are much less regulated than banks or insurance companies and can follow any trading strategy they like. To some extent they do add diversity (and therefore liquidity) to the market. But, as mentioned earlier, hedge funds tend to be highly leveraged. When liquidity tightens as it did in the second half of 2007, all hedge funds have to unwind positions accentuating the liquidity problems.

One conclusion from the arguments we have put forward is that a contrarian investment strategy has some merit. If markets overreact for the reasons we have mentioned, an investor can do quite well by buying when everyone else is selling and there is very little liquidity. However, it can be quite difficult for a financial institution to follow such a strategy if it is subject to short-term VaR-based risk management.

Summary

There are two types of liquidity risk: liquidity trading risk and liquidity funding risk. Liquidity trading risk is concerned with the ease with which positions in the trading book can be unwound. The liquidity trading risk of an asset depends on the nature of the asset, how much of the asset is to be traded, how quickly it is to be traded, and the economic environment. The credit crisis of 2007 emphasizes the importance of transparency. Assets that are not well defined or well understood are unlikely to trade in a

liquid market for long. The liquidity of an asset at a particular time can be measured as the cash bid–ask spread or as the proportional bid–ask spread. The latter is the difference between the bid and ask price divided by the average of the bid and ask price. The cost of unwinding a position in the asset is half of the bid–ask spread. Financial institutions should monitor the cost of unwinding the whole trading book in both normal market conditions and stressed market conditions.

A trader, when faced with the problem of unwinding a large position in an asset, has a trade-off between the bid–ask spread and market risk. Unwinding quickly leads to high bid–ask spreads, but low market risk. Unwinding slowly leads to lower bid–ask spreads, but more market risk. The optimal trading strategy depends on (a) the bid–ask spread as a function of the quantity traded in a day and (b) the probability distribution for daily changes in the asset price. For any particular unwind strategy, the trader can choose a confidence level and calculate the unwind cost that will not be exceeded with the confidence level. The unwind strategy that minimizes this cost can then be determined.

Liquidity funding risk management is concerned with being able to meet cash needs as they arise. It is important for a financial institution to forecast its cash needs in both normal market conditions and stressed market conditions to ensure that they can be met with almost total certainty. Cash needs depend on depositor withdrawals, drawdowns on lines of credit, guarantees that have been made, defaults by counterparties, and so on. Sources of cash are instruments that can be readily converted into cash, borrowings in the wholesale market, asset securitizations, new depositors, cash itself, and (as a last resort) borrowings from a central bank. In June 2008, bank regulators issued a list of 17 principles describing how banks should manage their liquidity and indicated that they would be monitoring the liquidity management procedures of banks more carefully in the future.

The most serious liquidity risks arise from what are sometimes termed liquidity black holes. These occur when all traders want to be on the same side of the market at the same time. This may be because they have similar positions and manage risks in similar ways. It may also be because they become irrationally exuberant, overexposing themselves to particular risks. What is needed is more diversity in the trading strategies followed by market participants. Traders who have long-term objectives should avoid allowing themselves to be influenced by the short-term overreaction of markets.

Further Reading

- Almgren, R., and N. Chriss. “Optimal Execution of Portfolio Transactions.” *Journal of Risk* 3 (Winter 2001): 5–39.
- Bangia, A., F. Diebold, T. Schuermann, and J. Stroughair. “Liquidity on the Outside.” *Risk* 12 (June 1999): 68–73.
- Bank for International Settlements. “Liquidity Risk Management and Supervisory Challenges,” February 2008.

- Bank for International Settlements. "Principles for Sound Liquidity Risk Management and Supervision," September 2008.
- Brunnermeier, M. K., and L. H. Pedersen. "Market Liquidity and Funding Liquidity." *Review of Financial Studies* 22, no. 6 (2009): 2201–2238.
- Brunnermeier, M. K., and L. H. Pedersen. "Predatory Trading." *Journal of Finance* 60, no. 4 (2005): 1825–1863.
- Clunie, J. *Predatory Trading and Crowded Exits: New Thinking on Market Volatility*. Petersfield, UK: Harriman House, 2010.
- Persaud, A. D., ed. *Liquidity Black Holes: Understanding, Quantifying and Managing Financial Liquidity Risk*. London: Risk Books, 1999.

Practice Questions and Problems (Answers at End of Book)

- 21.1 What was the transparency problem in the subprime crisis of 2007?
- 21.2 An asset is quoted bid 50, ask 55. What does this mean? What is the proportional bid–ask spread?
- 21.3 Suppose that an investor has shorted shares worth \$5,000 of Company A and bought shares worth \$3,000 of Company B. The proportional bid–ask spread for Company A is 0.01 and the proportional bid–ask spread for Company B is 0.02. What does it cost the investor to unwind the portfolio?
- 21.4 Suppose that in Problem 21.3 the bid–ask spreads for the two companies are normally distributed. For Company A the bid–ask spread has a mean of 0.01 and a standard deviation of 0.01. For Company B the bid–ask spread has a mean of 0.02 and a standard deviation of 0.03. What is the cost of unwinding that the investor is 95% confident will not be exceeded?
- 21.5 A trader wishes to unwind a position of 60 million units in an asset over 10 days. The cash bid–ask spread as a function of daily trading volume, q , is $a + be^{cq}$ where $a = 0.2$, $b = 0.1$, and $c = 0.08$ and q is measured in millions. The standard deviation of the price change per day is \$0.1. What is the optimal strategy that minimizes the 95% confidence level for the costs?
- 21.6 Explain the difference between the liquidity coverage ratio (LCR) and the net stable funding ratio (NSFR).
- 21.7 Why is it risky to rely on wholesale deposits for funding?
- 21.8 What was the nature of the funding risk problems of Ashanti Goldfields and Metallgesellschaft?
- 21.9 What is meant by (a) positive feedback trading and (b) negative feedback trading? Which is liable to lead to liquidity problems?
- 21.10 What is meant by liquidity-adjusted VaR?
- 21.11 Explain how liquidity black holes occur. How can regulation lead to liquidity black holes?
- 21.12 Why is it beneficial to the liquidity of markets for traders to follow diverse trading strategies?

Further Questions

- 21.13 Discuss whether hedge funds are good or bad for the liquidity of markets.
- 21.14 Suppose that a trader has bought some illiquid shares. In particular, the trader has 100 shares of A, which is bid \$50 and ask \$60, and 200 shares of B, which is bid \$25 and ask \$35. What are the proportional bid–ask spreads? What is the impact of the high bid–ask spreads on the amount it would cost the trader to unwind the portfolio? If the bid–ask spreads are normally distributed with mean \$10 and standard deviation \$3, what is the 99% worst-case cost of unwinding in the future as a percentage of the value of the portfolio?
- 21.15 A trader wishes to unwind a position of 200,000 units in an asset over eight days. The cash bid–ask spread, as a function of daily trading volume q , is $a + be^{cq}$ where $a = 0.2$, $b = 0.15$, and $c = 0.1$ and q is measured in thousands. The standard deviation of the price change per day is \$1.50. What is the optimal trading strategy for minimizing the 99% confidence level for the costs? What is the average time the trader waits before selling? How does this average time change as the confidence level changes?

Chapter 22

Model Risk Management

The use of models by financial institutions is growing fast. Models are employed for making credit decisions, liquidity management, the evaluation of credit exposures, the valuation of derivatives, the calculation of risk measures such as value at risk (VaR) and expected shortfall (ES), the management of portfolios of financial assets for clients, the assessment of capital adequacy, customer relationship management, fraud detection, the identification of money laundering, and so on. Complex models, with advanced analytic techniques such as machine learning, are being used to automate some activities previously carried out by humans. (See Chapter 29 for a discussion of machine learning.) It is now recognized that large financial institutions need a model risk management function to ensure that models are appropriate for their intended purpose and are used in the correct way.

Models are approximations to reality. The art of building a model is to capture the key aspects of reality for a particular application without allowing the model to become so complicated that it is almost impossible to use. Nearly always, a model relies on some assumptions about the phenomenon being modeled. It is important to understand those assumptions and know when they are no longer appropriate.

It used to be the case that models could be developed by a quant on a spreadsheet and used with minimal oversight. This is no longer the case. Models must now be well documented and reviewed periodically to determine whether they are still appropriate and are working as intended. When changes are made to models, approval is usually necessary. These aspects of model risk management are typically not popular with model builders, who tend to enjoy the creative aspects of model development and dislike spending time justifying their models and documenting them.

Regulators have, since the Global Financial Crisis, become less comfortable with the use of internal models to determine capital requirements. They have also become more concerned about the way models are used for other purposes within financial institutions. Model risk is a component of operational risk. Regulators want banks to have a system for managing the risks presented by models. Models must be developed, validated, and used in accordance with this system.

This chapter starts by reviewing regulatory guidance on model risk management. It describes the role of the model validation group and then moves on to compare the way in which models are used in finance with the way they are used by physicists and other scientists. After that, it focuses on the use of models for valuation and risk management and examines some of the lessons that can be learned from the mistakes made in that area.

22.1 Regulatory Guidance

In April 2011, the U.S. Board of Governors of the Federal Reserve System published SR 11-7, which provides guidance to banks on effective model risk management.¹ This document has proved to be very influential, and U.S. banks are expected to ensure that their practices are consistent with the principles outlined in the document. Bank supervisors elsewhere have indicated that they expect to see a similar systematic approach to understanding and managing model risk. If a bank supervisor feels that a bank's model risk management systems are inadequate, it may request the bank to hold additional capital.

SR 11-7 defines model risk as “the potential for adverse consequences from decisions based on incorrect or misused model outputs and reports.” Model risk occurs primarily for two reasons:

1. The model may have fundamental errors when viewed from the perspective of its objectives. These errors may concern the data it uses, the way calculations are done, numerical procedures, assumptions made, and so on.
2. The model may be used incorrectly or inappropriately. It is important to understand the limitations of models. A model may have been developed assuming a particular set of market conditions or a particular type of customer behavior. When these change, the model may no longer be appropriate.

Financial institutions should attempt to identify the sources of model risk and their magnitude. Complex models generally lead to more model risk than simple models. When models are linked or when several models are based on the same underlying set of assumptions, there is liable to be a greater degree of model risk. It is important to monitor model performance to determine as early as possible when a model is, for whatever reason, not performing as well as expected.

¹See Board of Governors of the Federal Reserve System, Office of the Comptroller of the Currency, “Supervisory Guidance on Model Risk Management,” SR 11-7, April 2011.

22.1.1 Model Development

When a model is developed, documentation is important. There are several reasons for this. If the model developer leaves the organization and there has been no good documentation, it might be very difficult for others to assume responsibility for the model, and, as a result, the work of the developer might be lost. Regulators frequently ask to see model documentation when they evaluate the procedures used by a financial institution. Also, model validation and periodic review are facilitated if the model has been well documented.

An important aspect of model documentation is a statement of the purpose of the model, the underlying theory, an evaluation of industry practice, and a review of published research. It is important that the documentation clearly explains the underlying mathematics, numerical procedures used, assumptions made, and so on. The documentation should be sufficiently detailed that someone who is new to the area can understand what has been done. As indicated earlier, model documentation will often be viewed as a boring chore by the model developer. But in practice it often proves to be a useful discipline, forcing the model developer to step back and think about his or her work in a way not done before. Sometimes changes to the model result.

Models should be tested as thoroughly as possible. This should include careful testing of all parts of the computer code and evaluating the model's behavior for a range of input data. The testing should identify situations where the model performs well and situations where it performs poorly. This can be done by observing the performance of the model with progressively more extreme values for the inputs. In some situations, it is appropriate to compare the output from the proposed model with the output from a more elaborate model that would be computationally too slow to be used in practice.

The way in which the performance of a model is tested will depend on the nature of the model. Models developed from historical data should be tested out of sample. For example, if a model for making loan decisions is developed from a particular set of data collected over a particular period, it would clearly not be appropriate to test it using the same data. Similarly, if a portfolio management strategy for a client is developed by analyzing the behavior of asset prices during a particular period, it should be tested using data from another period.

We have discussed in earlier chapters how back-testing can be used for VaR models. It can be used in other situations as well. Suppose a model based on a particular set of assumptions is used to price and hedge barrier options. It should be possible to determine how well it would have worked in the past. Would a user of the model have made a profit on average? If so, would the profit have been sufficiently great to compensate for the risks being taken?

As a model is used in practice, a financial institution should collect information on whether it is reflecting economic and business realities and working in the way intended. There is a danger that the existence of a model may lead to a change in behavior. For example, a model for managing customer relationships may lead to customers behaving differently; a model for detecting money laundering might work well until would-be

money launderers learn (perhaps by trial and error) how the model works and adjust their behavior accordingly. Even model users may find ways of adjusting inputs to a model or the way the model is used to make themselves look good.

22.1.2 Model Validation

One result of the increasing attention being devoted to model risk management is the formation of model validation groups. These groups are responsible for verifying that all the models used by the organization are performing as expected. They provide a check on the work of the model developer and critically review what has been produced and how the model is performing.

The person charged with validating a model should be independent of the model developer and the model user. Ensuring independence is critical to the success of the model risk management function. The reporting lines and incentives should be chosen to promote independence. The success of the model validation function should be judged by the objectivity of the model reviews, the issues raised, and the actions taken by management to address the issues. Corporate culture is important. If objective thinking and the challenging of decisions are not encouraged, the model validation process is likely to be ineffective. Regulators are likely to ask to see instances where models and their use have been appropriately changed as a result of validation.

Models are sometimes used informally in a small way prior to initial validation. However, before a model becomes part of a financial institution's way of operating it is important that a first validation be carried out. The extent and rigor of the validation should depend on the risks presented by the model. Consider the situation where a model has been developed to automate billions of dollars of loan decisions. Here it is clearly important that the validation be thorough. Even if the model is performing well and in accordance with its objectives, the model validation group might insist on a staged implementation with close monitoring to reduce model risk.

Model validation should continue on an ongoing basis after the model has been approved for use. This ensures that the impact of changes in markets and business practices on the performance of the model is monitored. SR 11-7 recommends that models be reviewed by the validation group at least annually. Users of the model (particularly if they are different from the developers) can provide input, and the validation group can determine whether the performance of the model using recent data is as good as it was on the data available when the model was first produced. The outcomes from using the model (e.g., credit decisions made, return on investments from using the model, fraud attempts thwarted) should be documented and compared with what was expected.

Business Snapshot 22.1 describes the London Whale story, which is an example of a situation where the model validation group was under pressure to approve a model. This was because the model provided justification for reducing VaR, and therefore regulatory capital, when a huge position in one product was hedged with an opposite position in a similar product. If the model validation group had more time, it would have carried out a more careful review of the model. It might have argued that the correlation between

BUSINESS SNAPSHOT 22.1**The London Whale**

The London Whale was a trader, Bruno Iksil, who worked for the Chief Investment Office (CIO) at JPMorgan Chase (JPM). His nickname arose from his huge positions in credit default swaps on CDX and iTraxx indices. (As explained in Section 17.4, these are products that provide protection against defaults for a portfolio of companies.)

The CIO's responsibility was to invest the bank's surplus cash. At the end of 2011, the bank's positions in the credit indices were moderately large and it was net long (i.e., it was buying more protection against defaults than it was selling). This net long position hedged the CIO's positions in fixed-income products. JPM wanted to reduce its risk-weighted assets to comply with the new Basel regulations. It also wanted to move toward a more neutral position in the credit indices because it felt that the economy was improving. Unwinding its existing positions in the indices would lead to a loss, partly because the positions had a negative mark-to-market value and partly because the positions were sufficiently large that unwinding them would tend to move the market against JPM. As a result, the CIO chose to hedge its net long position in credit indices by selling protection on different indices.

The CIO expected the long and short positions to offset each other. In fact, this did not happen. Market participants realized that someone was taking big positions. (JPM, like other dealers, tries to keep its positions secret. However, market participants realized that there was a "whale" in the market and it eventually became apparent that the whale was a trader for JPM.) Realizing that the positions would have to be unwound at some future time and that this would move the market, some market participants engaged in predatory trading, taking opposite positions to JPM. This moved the market against JPM and meant that the values of JPM's long and short positions were not as closely correlated as expected.

The CIO increased the size of its positions. This was partly for technical hedging reasons and partly to defend against predators. Its huge position was theoretically well balanced (i.e., models showed that its value should not be affected to any great extent by credit spread movements). It had long and short positions in CDX investment-grade (IG) indices and in the CDX high-yield (HY, non-investment-grade) indices. It sold large amounts of credit protection on the 10-year CDX IG 9 index and bought large amounts of protection on the five-year CDX IG 9 index. (This was an index created from 125 names in September 2007. Although not the current CDX IG index, it still traded reasonably actively; 121 of the original 125 names had not defaulted.) Prices did

not move in the way that was expected, and a loss of about \$6 billion was recognized by April/May 2012. Deferred bonuses to CIO employees were clawed back, and Jamie Dimon, JPM's chairman and CEO, took a pay cut for the year. Allegations that the value of the portfolio was misreported to shareholders were investigated by U.S. government agencies, and JPM ended up paying a fine of almost \$1 billion.

Why was the CIO allowed to take such huge risks? In January 2012, the CIO exceeded its VaR limit and the bank as whole exceeded its VaR limit as a result of the CIO's trading. However, a new VaR model was being developed. This model, which was finished in the second half of January and quickly reviewed and approved by the model validation group, reduced the VaR of the CIO's portfolio by 50% so that pre-January VaR limits were adhered to. JPM's January 2013 internal report on the London Whale losses (particularly the appendix) indicates that a number of mistakes were made. The quant who developed the model and the model validation group were under pressure to develop and approve the model. The quant reported to traders in the CIO, and the model validation group had time to carry out only limited back-testing of the model. In May 2012, the model was re-evaluated. A number of serious errors were discovered and its use was discontinued.

This story emphasizes the importance of the model validation function and why it needs to be independent.

the two products was assumed to be too large. It could also have made the point that hedges that work well for small positions do not always work well for large positions. The bank's use of the model for its large position had the potential to lead to predatory trading (see Chapter 21), rendering the historical comparisons of the prices of the products being traded less relevant.

SR 11-7 argues that comprehensive validation should have three core elements:

1. Evaluation of conceptual soundness, including developmental evidence
2. Ongoing monitoring, including process verification and benchmarking
3. Outcomes analysis, including back-testing

A sound developmental process should produce documented evidence in support of all model assumptions. There is a danger that the model developer has a bias toward a particular type of model—or even an interest in showing off his or her quant skills by developing an unnecessarily complex model. The validation work should ensure that this has not happened. Stability is an important property of models. If a small change in an input produces a large change in the output, the model may be unstable and

inappropriate for its intended use. Model validation should include model stress testing where the model is subjected to extreme inputs to test its range of applicability.

As already mentioned, ongoing monitoring is a core part of model validation. This includes ensuring that all changes made to a model are tested and documented. If the model user and the model developer are the same person (as is sometimes the case), there is a temptation for the model user to tweak the code to improve model performance. This is unsatisfactory and can lead to a situation where the model is nothing more than a collection of tweaks with no unifying theory. If a model is not performing as well as it could, a review should be undertaken and possibly a new model with better foundations developed. It is important for a model validation function to investigate situations where a model is not used because managers feel the output is not consistent with their experience and judgment. This can be an indication that improvements or a new approach is necessary.

Outcomes analysis, such as back-testing, is important for monitoring a model's performance. Some work in this area should be carried out by the model developer and documented. This work should be checked by the model validation group and continued on an ongoing basis. The reasons for any differences between expected outcomes and actual outcomes should be analyzed. For example, a particular liquidity management model might be found to work well in low-interest-rate environments, but less so when interest rates increase. This type of analysis can lead to model improvements.

22.1.3 Vendor Models

Not all models are developed internally. Some are purchased from model vendors. Sometimes the purpose of a model that has been purchased is to benchmark a similar model developed internally. In this case, the vendor model can be a useful validation tool. In other situations, the vendor model may be used because it is less expensive than developing a similar model internally. When this is the case, vendor models should be subject to the same validation as internal models. However, they present particular challenges for the model validation group, as the modeling expertise is external to the organization and some aspects of the model may be proprietary.

Financial institutions should have processes in place for selecting vendor models. The vendor should be required to provide documentation explaining the model design, testing results, and model limitations. Financial institutions should insist that vendors conduct ongoing performance monitoring and outcomes analysis, and that they provide the results of this to their clients.

The validation group should be responsible for approving the use of a vendor model. It may not have access to the computer code, but it can conduct extensive analysis on the results produced by the model for a variety of inputs. It should insist that the vendor provide full information about the data used to develop the model. There should be a contingency plan for a situation in which the vendor goes out of business or decides to stop supporting the model. One approach here is to insist that the computer code and

all other proprietary material be placed in escrow and made available to the client in circumstances in which the vendor is no longer able to provide support.

22.2 Models in Physics and Finance

Many individuals who were well trained in physics now work in finance developing models. Some models in finance are structurally similar to those in physics. For example, the differential equation that leads to the famous Black–Scholes–Merton option pricing model is the heat-exchange equation that has been used by physicists for many years. However, as Derman has pointed out, there is an important difference between the models of physics and the models of finance.² The models of physics describe physical processes and are highly accurate. By contrast, the models of finance describe the behavior of market variables, customers, borrowers, and so on. The phenomena being modeled are a result of the actions of human beings. As a result, the models are at best approximate descriptions of the phenomena being modeled.

One important difference between the models of physics and the models of finance concerns model parameters. The parameters of models in physics generally do not change. For example, the gravitational pull on the surface of the earth is always 32 feet per second per second. By contrast, parameters in finance models change daily. The volatility used to price an option might be 20% one day, 22% the next day, and 19% the following day. Indeed, the parameters used on a day are often chosen to be consistent with market prices, not to determine them.

22.3 Simple Models: Expensive Mistakes

Models do not have to be highly complex to lead to losses. Business Snapshot 22.2 provides a simple example of this. Kidder Peabody's computer system did not account correctly for funding costs when a linear product was traded. As a result, the system indicated that one of the company's traders was making a large profit when in fact he was taking a huge loss. The model used by Kidder Peabody should have been subject to independent validation. Even if the theoretical problem with the model was not picked up, the difference between model prices and market prices would have indicated to the model validation group that something was seriously wrong.

The Kidder Peabody modeling mistake would have been easy to identify if there was a culture within the organization that allowed traders to be challenged. Indeed, Barry Finer, risk manager for the government bond desk, did point out the difficulty of making large arbitrage profits from a market as efficient as the U.S. government bond market. But his concerns were dismissed out of hand.

²See E. Derman, *My Life as a Quant: Reflections on Physics and Finance* (Hoboken, NJ: John Wiley & Sons, 2004), and E. Derman, "Model Risk," *Risk* 9, no. 2 (May 1996): 139–145.

BUSINESS SNAPSHOT 22.2**Kidder Peabody's Embarrassing Mistake**

Investment banks have developed a way of creating a zero-coupon bond, called a *strip*, from a coupon-bearing Treasury bond by selling each of the cash flows underlying the coupon-bearing bond as a separate security. Joseph Jett, a trader working for Kidder Peabody, had a relatively simple trading strategy. He would buy strips and sell them in the forward market. The forward price of the strip was always greater than the spot price, so it appeared that he had found a money-making machine! In fact, the difference between the forward price and the spot price represents nothing more than the cost of funding the purchase of the strip. Suppose, for example, that the three-month interest rate is 4% per annum and the spot price of a strip is \$70. The three-month forward price of the strip is $70e^{0.04 \times 3/12} = \70.70 . This is because, for someone who wants to own the strip in three months, the forward contract saves \$0.70 of funding costs.

Kidder Peabody's computer system reported a profit on each of Jett's trades equal to the excess of the forward price over the spot price (\$0.70 in our example). By rolling his contracts forward, Jett was able to prevent the funding cost from accruing to him. The result was that the system reported a profit of \$100 million on Jett's trading (and Jett received a big bonus) when in fact there was a loss in the region of \$350 million. This shows that even large financial institutions can get relatively simple things wrong!

Sometimes the assumptions made by model developers are inaccurate but less easy to identify than in the Kidder Peabody case. An example here is provided by the interest rate swap market. A plain vanilla interest rate swap, such as the one described in Section 5.5.3, can be valued by assuming that forward interest rates will be realized. This is explained in Appendix D. For example, if the forward interest rate for the period between 2 and 2.5 years is 4.3% per annum, the swap can be correctly valued on the assumption that the floating rate that is exchanged for a fixed rate at the 2.5-year point is calculated using this rate. It is tempting to generalize from this and argue that any swap agreement to exchange cash flows in the future can be valued on the assumption that forward rates are realized. This is not so. Consider, for example, what is known as an *in-arrears swap*. In this instrument, the floating rate that is observed on a particular date is paid on that date, not one accrual period later as is the case for a plain vanilla swap. An in-arrears swap should be valued on the assumption that the realized interest rate equals the forward interest rate plus a "convexity adjustment." As indicated in Business Snapshot 22.3, financial institutions that did not understand this lost money in the mid-1990s.

BUSINESS SNAPSHOT 22.3

Exploiting the Weaknesses of a Competitor's Mode

An in-arrears swap is an interest rate swap where the floating interest rate is paid on the day it is observed, not one accrual period later. Whereas a plain vanilla swap is correctly valued by assuming that future rates will be today's forward rates, an in-arrears swap should be valued on the assumption that a future rate is today's forward interest rate plus a "convexity adjustment."

In the mid-1990s, sophisticated financial institutions understood the correct approach for valuing an in-arrears swap. Less sophisticated financial institutions used the naive "assume forward rates will be realized" approach. The result was that by choosing trades judiciously, sophisticated financial institutions were able to make substantial profits at the expense of their less sophisticated counterparties.

The derivatives business is one where traders do not hesitate to exploit the weaknesses of their competitors' models!

A similar example to this is provided by the interest rate futures markets. In most situations, it is reasonable to assume that a futures price is the same as the corresponding forward price. For example, the futures price in a futures contract to deliver gold in one year should be very close to the forward price in a forward contract to deliver gold at the same time. There are some differences between the two contracts. For example, a futures contract is settled daily and the party with the short position has some flexibility on the precise delivery date. But these differences do not create a material difference between the futures and forward prices. In the mid-1990s, many market participants chose to assume that the futures rate in an interest rate futures contract is the same as the corresponding forward rate. In fact, the approximate result that is true for most other futures contracts does not apply to interest rate futures. This is because interest rate futures last a long time so that the impact of daily settlement cannot be ignored. Also, in the case of interest rates, futures and forwards are settled at different times. Futures are settled at the beginning of the period covered by the underlying rate whereas forwards are settled at the end.³

22.3.1 Monitoring Trading Patterns

The situations we have just outlined suggest that an important way of identifying model risk in the valuation of products is to monitor trading patterns. In particular,

³See J. Hull, *Options, Futures, and Other Derivatives*, 11th edition, Pearson, 2022 for a more detailed discussion.

the risk management function within a financial institution should keep track of the following:

1. The type of trading the financial institution is doing with other financial institutions
2. How competitive it is in bidding for different types of structured transactions
3. The profits being recorded from the trading of different products

Getting too much of a certain type of business, or making huge profits from relatively simple trading strategies, can be a warning sign. If a financial institution finds that its prices are out of line with the market, it must make adjustments to its valuation procedures to bring them into line.

The high profits being recorded for Joseph Jett's trading at Kidder Peabody (see Business Snapshot 22.2) should have indicated that something was amiss. Likewise, if in the mid-1990s a financial institution's risk management team discovered that traders were entering into a large number of in-arrears swaps with other financial institutions (see Business Snapshot 22.3) where they were receiving fixed and paying floating, they could have alerted modelers to a potential problem and directed that trading in the product be temporarily stopped.

22.4 Models for Pricing Actively Traded Products

When a financial instrument trades actively, we do not need a model to know what its price is. The market tells us this. Suppose, for example, that a certain option on a stock index trades actively and is quoted by market makers as bid \$30, ask \$30.5. Our best estimate of its current value is the mid-market price of \$30.25.

A situation that is common in the over-the-counter market is one where a financial instrument that has to be valued is a standard product such as an option, but not exactly the same as one that trades in the market. For example, it might be an option with a strike price or time to maturity (or both) different from the options whose prices can be observed. The model is then used as a tool to ensure that the way an instrument is priced is consistent with the observed market prices of other similar instruments. A good example of how this is done is provided by the way the Black–Scholes–Merton model is used in practice. (See Appendix E for a description of the Black–Scholes–Merton model.)

The Black–Scholes–Merton model has been successful because all its inputs except one are known with a high degree of certainty. The one input that cannot be observed is volatility. This means that there is a one-to-one correspondence between prices and volatilities. When the volatility for an option has been specified, the price can be calculated. Similarly, when the price is known, a volatility can be determined. The volatility determined from a market price is known as an *implied volatility*. Traders responsible for options on a particular asset keep track of what is known as the *volatility surface*. This is the set of implied volatilities calculated for options with different strike prices and times to maturity.

Table 22.1 Volatility Surface

Volatilities for different strike prices and maturities are shown as % per annum.

	Strike Price				
	0.90	0.95	1.00	1.05	1.10
1-month	14.2	13.0	12.0	13.1	14.5
3-month	14.0	13.0	12.0	13.1	14.2
6-month	14.1	13.3	12.5	13.4	14.3
1-year	14.7	14.0	13.5	14.0	14.8
2-year	15.0	14.4	14.0	14.5	15.1
5-year	14.8	14.6	14.4	14.7	15.0

Table 22.1 provides an example of the volatility surface that could be constructed for options on an exchange rate. (For simplicity we assume that the current exchange rate is close to 1.00.) Note that, if the Black–Scholes–Merton model were a perfect description of reality, the volatilities in the table would be constant and not vary through time. In practice, implied volatilities for options are not the same for different strike prices and times to maturity, and they do vary through time. (This is true for options on all assets.) It is the trader's job to understand the current volatility surface and how it could change through time.

Suppose that a trader wishes to value a nine-month option with a strike price of 1.05 on the day Table 22.1 is produced. A six-month option with this strike price has a volatility of 13.4% and a one-year option with this strike price has a volatility of 14.0%. Interpolating between these, a trader would use a volatility of 13.7% in the Black–Scholes–Merton model to value the option. Suppose next that both the time to maturity and the strike price are different from those in the table. A two-dimensional interpolation is then necessary. This involves first interpolating between strike prices and then between times to maturity. (The same answer is obtained if one first interpolates between maturities and then between strike prices. See Problem 22.12.) Consider a 1.5-year option with a strike price of 0.92. A one-year option with this strike price has an interpolated implied volatility of 14.42%; a two-year option with the strike price has an interpolated volatility of 14.76%. The interpolated volatility for a 1.5-year option is therefore 14.59%.

22.4.1 Sources of Model Risk for Actively Traded Products

One issue for the model validation group is whether the prices determined for actively traded instruments depend on the model used. If, in the example just considered, we replaced the Black–Scholes–Merton model by another model, would the prices be similar? The answer to this question is that, when the nature of the product being valued is exactly the same as actively traded instruments (i.e., only parameters like the strike price

and time to maturity are different), the model does not usually affect prices materially.⁴ This is because the model is being used as nothing more than a sophisticated interpolation tool and two different interpolation tools usually give very similar valuations.

There is liable to be model risk in extreme situations. The model validation group should check this out. For example, it should examine the operation of different models when the volatility surface is very steep in either the strike price or time-to-maturity dimension. Also, the model validation group should check on the reliability of the pricing data that are available in the market and how the model would work when less data than usual are available.

22.4.2 Hedging

In the case of actively traded instruments, the main model risk concerns hedging. As we have discussed, switching from one model to another does not usually affect prices to any great extent, but it can have an effect on hedging performance.

We can distinguish between *within-model hedging* and *outside-model hedging*. Within-model hedging is designed to deal with the risk of changes in variables that are assumed to be uncertain by the model. Outside-model hedging deals with the risk of changes in variables that are assumed to be constant (or deterministic) by the model. For the Black–Scholes–Merton model, hedging against movements in the underlying stock price (delta and gamma hedging) is within-model hedging, because the model assumes that stock price changes are uncertain. However, hedging against volatility changes (vega hedging) is outside-model hedging because the model assumes that volatility is constant.

In practice, traders almost invariably do outside-model hedging as well as within-model hedging. This is because, as we have explained, the calibration process results in parameters such as volatilities (which are assumed by a model to be constant) changing daily. A good options trader will monitor the exposure of a trading book to the sorts of shifts in the volatility surface that are typically seen.⁵

If a model is designed to be always consistent with market prices and hedging is carried out continuously for all variables (both those that are assumed to be constant by the model and those that are assumed to be stochastic), the value of the hedger's position should in theory not change. However, in practice hedging cannot be carried out in this idealized way. Consider again the Black–Scholes–Merton model to value a position in options on an exchange rate. Each option price is a deterministic function of the exchange rate, its implied volatility, and time.⁶ This means that the trader has an exposure to the exchange rate and the implied volatilities of all the options in the portfolio. The exposure to the exchange rate can be hedged without difficulty by trading the foreign

⁴This is demonstrated by S. Figlewski, "Assessing the Incremental Value of Option Pricing Theory Relative to an Informationally Passive Benchmark," *Journal of Derivatives* (Fall 2002): 80–96.

⁵A principal components analysis, as described in Chapter 14, can be used. If exposure to each of the major principal components is small, the exposure to the shifts that have been observed in practice should also be small.

⁶To keep this example simple, it is assumed that interest rates do not change.

currency. This is known as delta hedging and is usually done at least once a day (see Chapter 15). Hedging the risk of volatility changes is not as easy. It requires trades in other options (which can be expensive) and assumptions about how the volatility surface can change through time.⁷ As a result, volatility hedging cannot be anything near as effective as delta hedging.

In practice, delta hedging is often the only hedging that a trader carries out on a regular basis. The delta of an option position does depend on the model used. This means that the hedge position taken by a trader varies from one model to the next. An important activity for the model validation group is to test the effectiveness of delta hedging for alternative models.

22.4.3 *P&L Decomposition*

One analysis that can be carried out is known as *P&L decomposition*. Risk managers divide day-to-day changes in the value of a portfolio into:

1. A change resulting from risks that were unhedged
2. A change resulting from the hedging model being imperfect
3. A change resulting from new trades done during the day

If the day-to-day change is unacceptable, the analysis will indicate the areas where more effort should be expended. The good news here is that, on average, the total gain or loss to a financial institution from inaccurate hedging is often small because it is largely diversified away across the portfolio of a large financial institution.

22.5 Models for Less Actively Traded Products

The general procedure for pricing derivatives is to build a model and then imply the parameters of the model from the prices that can be observed in the market. This is illustrated in Figure 22.1. One or more other financial instruments whose prices can be observed in the market are chosen as the “calibrating instruments.” These instruments are chosen so that they are as similar as possible to the financial instrument of interest. They are then used to imply parameters for the model being used for the instrument under consideration.

We have argued that there is very little model risk as far as valuation is concerned when this approach is used for standard instruments that trade actively. The same is not true for non-standard instruments. Here the model used is likely to have a significant effect on valuations as well as on hedging. An example of model risk in the pricing of structured products is provided by Business Snapshot 7.1. This makes the point that the

⁷If perfect hedging against volatility surface movements were possible, differences in delta hedging would be offset by differences in volatility hedging. But volatility hedging is in practice far from perfect.

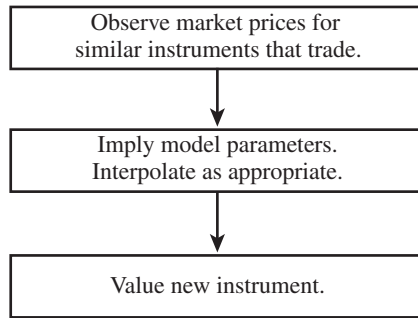


Figure 22.1 How Models Are Usually Used for Valuing a Financial Instrument

assumption that the BBB-rated tranche of an asset-backed security (ABS) is similar to a BBB bond is not a good one when tranches of an ABS CDO are being priced.

An important point to note is that we are nearly always concerned with what might be termed *relative valuation*. The objective is to value the financial instrument consistently with other similar financial instruments whose prices can be observed. It is very rare to use a model to value a derivative in absolute terms. This is another difference between finance models and the models of physics.

A financial institution should not rely on a single model for pricing non-standard products. Instead it should, whenever possible, use several different models. This leads to a price range for the instrument and a better understanding of the model risks being taken.

Suppose that three different models give prices of \$6 million, \$7.5 million, and \$8.5 million for a particular product that a financial institution is planning to sell to a client. Even if the financial institution believes that the first model is the best one and plans to use that model as its standard model for daily repricing and hedging, it should ensure that the price it charges the client is at least \$8.5 million. Moreover, it should be conservative about recognizing profits. If the product is sold for \$9 million, it is tempting to recognize an immediate profit of \$3 million (\$9 million less the believed-to-be-accurate price of \$6 million). However, this is overly aggressive. A better, more conservative, practice is to put the \$3 million into a reserve account and transfer it to profits slowly during the life of the product.⁸

22.5.1 Weighted Monte Carlo Simulation

When different models are being used, they should be calibrated to the prices of actively traded standard products that are observed in the market. For example, when a barrier option is being priced, the models tested should produce prices for regular call and put options that are equal to those observed in the market. Rather than testing a small number of different models, a sophisticated analysis can be attempted to answer the question: "What range of model prices is possible for models that price actively

⁸This is also likely to have sensible implications for the way bonuses are paid.

traded products correctly?” This analysis, if carried out successfully, will give a worst possible price and a best possible price for the non-standard instrument. One approach is known as *weighted Monte Carlo simulation*.⁹ This involves carrying out a Monte Carlo simulation for one model and then applying weights to the sampled paths. Constraints are placed on the weights so that standard instruments are priced correctly by the weighted Monte Carlo simulation. An optimization procedure is used to find the weights that produce maximum and minimum values for the non-standard derivative consistent with the constraints.

22.6 Accounting

Products that are held in the trading book must be valued daily. This process is referred to by traders as *marking to market*. The valuations are used by accountants to calculate quarterly financial statements. Accountants refer to marking to market as “fair-value accounting.”

Accountants recognize that some models produce more reliable prices than others. The Financial Accounting Standards Board in the United States (FAS 157) and the International Accounting Standards Board in Europe (IAS 39) require banks to report three categories of valuations separately:

Level 1 valuations: Valuations for instruments where there are quoted prices in active markets.

Level 2 valuations: Valuations for instruments where there are quoted prices for similar instruments in active markets or quoted prices for the same instrument in markets that are not active.

Level 3 valuations: Valuations for instruments where some modeling assumptions are required. (This is sometimes referred to as *marking to model*.)

The valuations we discussed in Section 22.4 where the model is used as a sophisticated interpolation tool would be classified as Level 2 valuations. The valuations in Section 22.5 where the calibrating instruments are somewhat different from the instrument being valued would be classified as Level 3 valuations.

22.7 What Makes a Successful Pricing Model?

It might be argued that any model where the values of input parameters change from day to day is not helpful. Certainly this would be true for models in physics. But, as has been explained, models for valuing financial instruments are used in a quite different way from the models of physics.

The Black–Scholes–Merton model is popular for three reasons:

⁹See, for example, A. Elices and E. Gimenez, “Weighted Monte Carlo,” *Risk* (May 2006).

1. It can be used in the way we have described for interpolation in conjunction with a volatility surface, and ensures that prices are consistent with observed market prices.
2. It is a communication tool. Traders often prefer quoting implied volatilities to quoting option prices, because implied volatilities are more stable than option prices and, as a result, their quotes do not change as frequently. When the price of the underlying asset changes or the interest rate changes, the option price changes but the implied volatility may not.
3. The model is sufficiently simple that a trader can develop intuition about the model and use it to structure his or her thinking about option markets.

Models in other areas that have these properties have also proved to be popular. Model developers and the model validation group should understand that the ultimate purpose of a model is often to enable the model user to understand the market and exercise his or her judgment more efficiently. When alternative models are compared, this is an important consideration and may lead to simpler models being preferred to more complex ones.

22.8 Model-Building Missteps

The art of model building is to capture what is important for valuing and hedging an instrument without making the model more complex than it needs to be. Sometimes models have to be quite complex to capture the important features of a situation, but this is not always the case.

One potential misstep in model building is *over-fitting*. Consider the problem posed by the volatility surface in Table 22.1. We can exactly match the volatility surface with a single model by extending the Black–Scholes–Merton model so that volatility is a complex function of the underlying asset price and time.¹⁰ This can be desirable for some applications, but the model can become difficult to use. We may find that some properties of the model are less reasonable than those of simpler models. In particular, the joint probability distribution of the asset price at two or more times might be less than ideal, and, in some circumstances, future volatility surfaces given by the model might look quite different from those observed in the market today.¹¹

¹⁰This is the implied volatility function model proposed by B. Dupire, “Pricing with a Smile,” *Risk* 7 (February 1994): 18–20; E. Derman and I. Kani, “Riding on a Smile,” *Risk* (February 1994): 32–39; M. Rubinstein, “Implied Binomial Trees,” *Journal of Finance* 49, no. 3 (July 1994): 771–818.

¹¹Models that fit the volatility surface at all future times accurately price instruments where there is a payoff at one future time that is dependent only on the asset price at that time. However, they are liable to be less accurate for instruments such as barrier options and compound options that depend on the joint probability distribution of the asset price at two or more times. Hull and Suo find that the implied volatility function model works reasonably well for compound options, but is less accurate for barrier options. See J. Hull and W. Suo, “A Methodology for the Assessment of Model Risk and Its Application to the Implied Volatility Function Model,” *Journal of Financial and Quantitative Analysis* 37, no. 2 (June 2002): 297–318.

Another potential misstep in model building is *over-parameterization*. The Black–Scholes–Merton model can be extended to include features such as stochastic volatility or jumps in the asset price. This invariably introduces extra parameters that have to be estimated and makes the model more difficult to use. It is usually claimed that the parameters in complex models are more stable than those in simpler models and do not have to be adjusted very much from day to day. This may be true, but we should remember that we are not dealing with physical processes. The parameters in a complex model may remain relatively constant for a period of time and then change, perhaps because there has been what economists refer to as a *regime change*. A financial institution may find that a more complicated model is an improvement over a simple model until there is a regime change. The more complicated model may not then have the flexibility to cope with changing market conditions.¹²

As we have already mentioned, simple models are often the most successful. Often users are skeptical of complex models because they are so-called black boxes and it is very difficult to develop intuition about them. In some situations, users' skepticism is well founded for the reasons we have just mentioned.

Summary

Model risk is a form of operational risk. It is now recognized that it requires special attention. Regulators expect to see that a financial institution has comprehensive procedures for developing, documenting, validating, and using models. Models should be reviewed periodically to ensure that they are working as originally intended. The model validation group should be independent of model developers and model users. It should provide a check on the work of the model developer and reach conclusions about the circumstances under which the model should be used.

Models in finance are different from those used in the physical sciences, because they are ultimately models of human behavior. They can usually only provide an approximation to reality. Furthermore, from time to time there are regime shifts leading to fundamental shifts in the performance of models.

Valuation models are often nothing more than sophisticated interpolation tools to ensure that one instrument is priced consistently with other similar instruments. As the instrument being valued becomes less standard, the valuation model chosen becomes more important. It is then a good practice for the model validation group to test the use of different models to obtain a realistic range for pricing and to understand the accompanying model risk. The models should always be calibrated to the current prices of actively traded instruments. The hedging performance of a model, as well as valuation performance, should be evaluated carefully.

¹²The nature of this type of problem in the social sciences is discussed in the famous Lucas Critique, which was propounded by the economist Robert Lucas in 1976. See R. Lucas, "Economic Policy Evaluation: A Critique," *Carnegie-Rochester Conference Series on Public Policy* 1 (1976): 19–46.

Further Reading

- Avellaneda, M., R. Buff, C. Friedman, N. Grandchamp, L. Kruk, and J. Newman. "Weighted Monte Carlo: A New Technique for Calibrating Asset Pricing Models." *International Journal of Theoretical and Applied Finance* 4, no. 1 (February 2001): 91–119.
- Bates, D. S. "Post '87 Crash Fears in the S&P Futures Market." *Journal of Econometrics* 94 (January/February 2000): 181–238.
- Board of Governors of the Federal Reserve System, Office of the Comptroller of the Currency. "Supervisory Guidance on Model Risk Management," April 2011.
- Cont, R. "Model Uncertainty and Its Impact on the Pricing of Derivative Instruments." *Mathematical Finance* 16, no. 3 (July 2006): 519–547.
- Derman, E. "Model Risk," *Risk* 9, no. 5 (May 1996): 139–145.
- Derman, E. *My Life as a Quant: Reflections on Physics and Finance*. Hoboken, NJ: John Wiley & Sons, 2004.
- Elices, A., and E. Gimenez. "Weighted Monte Carlo." *Risk* (May 2006).
- Glasserman, P., and X. Xu. "Robust Risk Measurement and Model Risk." *Quantitative Finance* 14, no. 1 (January 2014): 29–58.
- Hull, J. C., and W. Suo. "A Methodology for the Assessment of Model Risk and Its Application to the Implied Volatility Function Model." *Journal of Financial and Quantitative Analysis* 37, no. 2 (June 2002): 297–318.
- JPMorgan Chase. "Report of JPMorgan Chase & Co. Management Task Force Regarding 2012 CIO Losses," January 2013.

Practice Questions and Problems (Answers at End of Book)

- 22.1 Explain what is meant by (a) marking to market and (b) marking to model.
- 22.2 What are the two categories of model risk identified in SR 11-7?
- 22.3 Why is it important that a model be fully documented?
- 22.4 "The Black–Scholes–Merton model is nothing more than a sophisticated interpolation tool." Discuss this viewpoint.
- 22.5 Using Table 22.1, calculate the volatility a trader would use for an eight-month option with a strike price of 1.04.
- 22.6 What is the key difference between the models of physics and the models of finance?
- 22.7 How is a financial institution likely to find that it is using a model different from its competitors for valuing a particular type of derivatives product?
- 22.8 Distinguish between within-model and outside-model hedging.
- 22.9 What is meant by P&L decomposition?
- 22.10 What is meant by "Level 1," "Level 2," and "Level 3" valuations in accounting?
- 22.11 What is meant by over-fitting and over-parameterization in model building?
- 22.12 Section 22.4 calculates the implied volatility of an option with strike price 0.92 and time to maturity 1.5 years using the data in Table 22.1 as 14.59%. It does this by interpolating first along the strike price dimension and then along the time to

maturity dimension. Show that the same answer is obtained by interpolating first along the time to maturity dimension and then along the strike price dimension.

- 22.13 “For structured products, traders mark to model. They do not mark to market.” Explain this remark.

Further Questions

- 22.14 Suppose that all options traders decide to switch from Black–Scholes–Merton to another model that makes different assumptions about the behavior of asset prices. What effect do you think this would have on (a) the pricing of standard options and (b) the hedging of standard options?
- 22.15 Using Table 22.1, calculate the volatility a trader would use for an 11-month option with a strike price of 0.98.
- 22.16 Suppose that a financial institution uses an imprecise model for pricing and hedging a particular type of structured product. Discuss how, if at all, it is likely to realize its mistake.

Chapter 23

Climate Risk, ESG, and Sustainability

We now move on to consider environmental risks. These are the risks that we will damage the environment so much that the Earth will become less habitable (or even totally uninhabitable) for future generations. All companies, including financial institutions, are finding that, as a result of societal and other pressures, environmental risks can no longer be ignored. It seems almost certain that these risks will become increasingly important in the years ahead. The biggest environmental risk is climate risk, and this is what the first part of this chapter focuses on.

23.1 Climate Risk

The inhabitants of planet Earth add tens of billions of tons of greenhouse gases to the atmosphere each year. This is the major cause of global warming. When the energy from the sun reaches the Earth, some is absorbed and some is radiated back as heat. Increases in greenhouse gas concentrations in the atmosphere lead to less heat finding its way into space and more being absorbed by the atmosphere and the oceans. Unfortunately, this problem is not naturally self-correcting, as greenhouse gases stay in the atmosphere for hundreds of years.

The most prevalent greenhouse gas is carbon dioxide (CO₂).¹ As is well known, this is emitted by the burning of fossil fuels such as coal, oil, and natural gas. But it also arises from many other activities such as manufacturing cement, concrete, steel, and plastics, and raising animals for food. Achieving a reduction in greenhouse gas emissions is now recognized as essential for the survival of the human race. The Paris Agreement in 2015 had a goal of limiting temperature rises above preindustrial levels during the twenty-first century to 1.5 degrees Celsius. The Glasgow Climate Pact in 2021 reaffirmed this objective and strengthened targets for the year 2030 with a goal of reaching net-zero emissions by the year 2050. It is expected that the progress will be reviewed at further meetings between countries at intervals of about five years.

How successful we will be in achieving the goals set in Paris and Glasgow remains to be seen. Arguably, a temperature rise of less than 1.5 degrees Celsius is optimistic. The actual rise during the twenty-first century could be as high as 4 degrees Celsius or even 6 degrees Celsius. The primary reason for the broad range of estimates is not the physical science but the uncertainty around the global policy response and its success in actually reducing emissions. A number of organizations have developed models and produced climate scenarios. These organizations include the Intergovernmental Panel on Climate Change (IPCC), the Network for Greening the Financial System (NGFS), and the International Energy Agency (IEA). The IPCC calls these scenarios Shared Socioeconomic Pathways (SSPs) and published its sixth assessment report in 2021 (Table 23.1).²

The key problem is that humans have to make a trade-off between the emission of greenhouse gas today and the resultant damage to the environment in the future, which will impact their children and grandchildren. There is no easy way to make this trade-off. Table 23.2 shows 2020 estimates of greenhouse gas emissions by sector, produced by

Table 23.1 Shared Socioeconomic Pathways in IPCC Sixth Assessment Report

SSP	Scenario	Estimated Warming, 2041–2060, °C	Estimated Warming, 2081–2100, °C	Very Likely Range ³ 2081–2100, °C
SSP1	CO ₂ emissions cut to net zero by about 2050	1.6	1.4	1.0–1.8
SSP2	CO ₂ emissions cut to net zero by about 2075	1.7	1.8	1.3–2.4
SSP3	CO ₂ emissions continue at about current level until 2050 and then fall	2.0	2.7	2.1–3.5
SSP4	CO ₂ emissions double by 2100	2.1	3.6	2.8–4.6
SSP5	CO ₂ emissions triple by 2075	2.4	4.4	3.3–5.7

¹Other greenhouse gases that are created by human activities include methane, nitrous oxide, halo-carbons, and ozone.

²See IPCC Sixth Assessment Report, 2021, www.ipcc.ch/assessment-report/ar6.

³This can be interpreted as a 95% confidence interval.

Table 23.2 Greenhouse Gas Emissions by Economic Sector in the United States in 2020

Sector	Emissions
Transportation	27%
Electricity	25%
Industry	24%
Commercial and residential	13%
Agriculture	11%

the U.S. Environmental Protection Agency.⁴ Eliminating greenhouse gas emissions will require many changes to the way we produce energy, transport ourselves, manufacture materials, and feed ourselves. It will not be easy for developed countries. Furthermore, it may be an unreasonable burden for developing countries, as it is likely to adversely affect their economic progress. China is responsible for the most greenhouse gas emissions (about 10 billion tons per year) followed by the United States (about 5 billion tons per year) and India (about 2.5 billion tons per year). But on a per capita basis, the United States emits about twice as much as China and seven times as much as India.⁵

Climate change is going to be accompanied by rising sea levels as ice melts and sea water expands. It has been estimated that the sea level will rise a few meters for each degree Celsius of global warming.⁶ Some CO₂ will be absorbed by oceans, creating ocean acidification that adversely affects ocean ecosystems. Precipitation is expected to increase by about 3% for every degree Celsius of global warming because of evaporation from oceans.⁷ However, this will not be spread evenly. Areas that already get a lot of rain will get more while areas that are dry will become drier. As a result, floods can be expected in some parts of the world. Other extreme weather events such as storms, heat waves, wildfires, hurricanes, and tornadoes can also be expected. Indeed, we are already seeing them, and their costs total billions of dollars. In 2021, it is estimated that Hurricane Ida in the United States cost \$75 billion, flooding in China cost \$30 billion, and flooding in Germany and Belgium cost \$22 billion.⁸ The impact on human life is liable to be severe. In 2003, a heat wave in Europe caused tens of thousands of deaths. In 2022, parts of the world experienced an even worse heat wave.

⁴See EPA, “Sources of Greenhouse Gas Emissions,” www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions.

⁵See www.worldometers.info/co2-emissions/co2-emissions-by-country for statistics on all countries.

⁶See J. Garbe, T. Albrecht, A. Levermann, J. F. Donges, and R. Winkelmann, “The Hysteresis of the Antarctic Ice Sheet,” *Nature* 585 (September 2020): 538–544.

⁷See N. Jeevanjee and D. M. Roms, “Mean Precipitation Change from a Deepening Troposphere,” *Proceedings of the National Academy of Sciences* 115 (45), (October 2018): 11465–11470.

⁸See Jeff Masters, “The Top 10 Climate Change Events of 2021,” Yale Climate Connection (January 11, 2022), <https://yaleclimateconnections.org/2022/01/the-top-10-global-weather-and-climate-change-events-of-2021/>.

In this chapter, we consider the effect of climate risk on financial institutions. There are many uncertainties. Possibly, technological innovations (e.g., a cost-efficient way of removing CO₂ from the atmosphere and storing it permanently, or a cost-efficient way of capturing CO₂ before it is released into the atmosphere and storing it permanently will happen). But it is also possible that international agreements prove to be ineffective. There could be a ripple effect. If one or two large countries continue to be high emitters, a few other countries might decide not to follow through on their plans to reduce emissions because it will make little overall difference and could disadvantage them economically. This, in turn, could lead to yet more countries abandoning their green initiatives, and so on. This is a concern because there has already been a tendency for large countries to make climate commitments and then not follow through on them.

There have been some encouraging developments. One has been the establishment of the Task Force on Climate-Related Financial Disclosures (TCFD) by the Financial Stability Board. This develops recommendations on information related to climate risks that companies should disclose. Led by Michael Bloomberg, TCFD was supported by 3,000 organizations in 92 countries with a combined market capitalization of \$27.2 trillion in 2022. As we discuss later, Mark Carney's Glasgow Financial Alliance for Net Zero (GFANZ) has lined up vast amounts of investment capital that are committed toward transitioning to a low carbon future.

Analysts distinguish *physical risks* from *transition risks*. Physical risks are the costs of extreme climate events such as floods, rising sea levels, hurricanes, and so on. Transition risks are the risks associated with the move to lower emissions. These risks include the introduction of carbon taxes and costs incurred to reduce emissions by changing business practices or discontinuing certain activities altogether. For example, companies may find it necessary to close down plants that produce energy from fossil fuels earlier than planned. Business Snapshot 23.1 provides an example of this. However, the transition to lower emissions may give rise to opportunities for some companies as they find environmentally friendly ways to produce electricity, cement, and other products that currently lead to the production of greenhouse gases.

Other relevant risks related to climate change are those associated with large capital flows that may alter financial markets. GFANZ estimates that transitioning the economy will require additional investments on the order of \$100–150 trillion over the next 30 years. Finally, there are the risks of individual countries taking somewhat radical steps to mitigate the effects of climate change. For example, Solar Radiation Management, a form of geo-engineering that increases the ability of the atmosphere to reflect sunlight, is a short-term response that could have a temporary cooling effect that is both noticeable as well as quite affordable (only a few billion dollars).⁹ However, the consequences of such actions are very uncertain and would likely affect ocean currents, create droughts, or rainfall events in neighboring countries, as well as having biological impacts. In spite

⁹See, for example, Simon Nicholson, "Solar Radiation Management," Wilson Center, September 30, 2020, <https://www.wilsoncenter.org/article/solar-radiation-management>.

BUSINESS SNAPSHOT 23.1**Transition Costs in Germany**

In July 2020, Germany passed a law requiring that all coal-fired plants be closed by 2038 with some compensation being provided to owners. This will lead to many plants being closed down before the end of their useful lives (typically 30–50 years). For example, a coal power station, Datteln, became operational in 2020 and would have a maximum life of only 18 years if closed as planned. There will be a cost to German taxpayers.

of this, there are early pilot programs underway in an attempt to delay the effects of climate change for the most affected countries.

Governments have introduced some measures to reduce CO₂ emissions. About 25 countries have introduced carbon taxes. Sweden was the first country to do so in 1991 and has increased the tax rate per ton of fossil CO₂ emitted in 1997 and 2007. According to studies that have been carried out, the tax has been instrumental in reducing emissions. In the United States, a carbon tax has been repeatedly proposed but not implemented. Another measure is referred to as *cap and trade*. The amount of CO₂ that a company emits is capped, but companies can trade with each other. If a company does not need to emit its full quota, it can sell the unused part to another company that wants to emit more than its quota. This can lead to emissions being reduced where it is easiest and cheapest to do so. Cap and trade schemes are used in many jurisdictions, including China and some U.S. states.

23.1.1 Insurance Companies

Climate changes are likely to have a big impact on the insurance industry. Ideally, insurance companies would like to take on risks that are uncorrelated with each other. Consider the car insurance portfolio of a large insurance company. The claims made by different policy holders have very little correlation with each other. This means that a “law of large numbers” argument applies and the total payout in a year can be predicted reasonably accurately. We made the point in Chapter 3 that the same is not true for all of the risks the insurance company takes on. A climate-related disaster such as Hurricane Ida leads to many policy holders making claims at the same time. It seems almost certain that these types of natural disasters will occur more frequently in the future.

Insurance companies have over the years developed a great deal of expertise in modeling catastrophic risks. Their models must now recognize that the frequency and severity of these risks are increasing.¹⁰ The actuarial practice of averaging the cost of past risks

¹⁰See The Geneva Association, “Climate Change Risk Assessment for the Insurance Industry,” 2021. https://www.genevaassociation.org/sites/default/files/research-topics-document-type/pdf_public/climate_risk_web_final_250221.pdf.

to estimate the cost of future risks is not valid. Fortunately, the vast majority of property and casualty insurance and reinsurance contracts are for one year. This makes it possible to adjust the pricing, terms and conditions, and product offerings as climate conditions worsen.

Unlike property and casualty companies, life insurance companies do not have the flexibility of changing their premiums as conditions worsen. The annual premiums for whole life policies usually remain fixed for the life of the policy holder. As explained in Chapter 3, life expectancy in the United States has increased by about 20 years over the last 100 years. This trend has benefited life insurance companies because it means that premiums are received for longer and payouts are later than expected. It is possible that temperature increases and other disruptions will lead to a reversal of the life expectancy trends. Whole life policies might then become less profitable.

There can be no doubt that climate risks will make the management of insurance companies more challenging. Insurance companies must devote resources to understanding climate trends and quantifying their effect on risks they are underwriting. They need to convince regulators that they are handling climate risks in a responsible way. It is clearly in their interests and in the interests of society that policy holders be educated on ways of managing climate risks.¹¹ But an unfortunate reality is that insurance companies are becoming increasingly unwilling to provide insurance for natural disasters, such as floods, in areas they consider to be high risk. An important question for society is likely to be who bears the cost of damage to properties from natural disasters. Is it the property owners, private insurance companies, a government-sponsored insurance company, or disaster relief programs?

23.1.2 *Banks*

Banks are being proactive in trying to quantify the effects of climate change. The Partnership for Carbon Accounting Financials (PCAF) is a forum where banks exchange ideas on the best way to measure and address carbon emissions. The risks for banks arising from climate changes are also considered by their regulators. The European Central Bank in 2022 launched a supervisory climate stress test to assess how prepared banks are for dealing with financial and economic shocks arising from climate risk.¹² In a 2021 document, the Basel Committee on Banking Supervision argued that the impact of climate changes can be understood by considering their effect on traditional risk categories. Table 23.3 summarizes the Basel Committee's conclusions.¹³

¹¹For example, there are a number of simple steps home owners can take to avoid basement flooding. See https://www.intactcentreclimateadaptation.ca/wp-content/uploads/2021/03/3-Steps-to-Home-Flood-Protection_March-2021_Space-for-Partner-Logo.pdf?utm_source=kitchenertoday.com&utm_campaign=kitchenertoday.com&utm_medium=referral.

¹²See European Central Bank, "SCU Banking Supervision Launches 2022 Climate Risk Stress Test," Press release, January 27, 2022, bankingsupervision.europa.eu/ECB-Banking-Supervision-launches-2022-climate-risk-stress-test.pdf.

¹³See Basel Committee on Banking Supervision, "Climate-related risk drivers and their transmission channels." Bank for International Settlements, April 2021, www.bis.org/bcbs/publ/d517.pdf.

Table 23.3 Potential Impact of Climate Risks on Banks

Risk	Potential Effects of Physical and Transition Climate Risks
Credit risk	Climate risk may reduce borrowers' ability to service debt or the value of collateral in the event of a default.
Market risk	Prices of financial assets may suffer as the impact of climate risk becomes more apparent.
Liquidity risk	A bank's access to stable sources of funding such as retail deposits could be reduced as market conditions change.
Operational risk	There are likely to be legal and regulatory compliance risks associated with corporate clients and investments.
Reputational risk	Reputational risk may increase as a result of market or consumer sentiment.

Climate changes affect the risks associated with a bank's mortgage portfolio. As already pointed out, premiums can be changed each year in property and casualty insurance. By contrast, a mortgage entails a commitment over several years. For example, a mortgage in the United States typically lasts 15 or 30 years.

The occurrence of climate-related events, or the market at some future time becoming more aware of the possibility of these events, can affect real estate prices and therefore the value of mortgage collateral. As we saw during the Global Financial Crisis (see Chapter 7), when the value of a house drops below the amount outstanding on the mortgage, the homeowner may have an incentive to default even if he or she has the financial resources to continue to service the mortgage. A default leads to the bank taking possession of the house and having to sell it in a difficult market where other similar properties are being sold for similar reasons. In some jurisdictions, the bank has no recourse to other assets of the borrower.

It is not difficult to imagine climate risks causing similar problems to those created by the Global Financial Crisis. The possibility of rising sea levels, for example, could become an increasing concern in certain areas, reducing house prices and leading owners to bail out by defaulting on mortgages. When valuing a house for the purposes of determining the amount that can be safely lent, it is clearly important that banks take account of the potential impact of adverse climate events.

The evidence is that severe climate-related events can reduce corporate profitability and therefore increase credit risk to their lenders.¹⁴ Supply chains can be disrupted, worker productivity can be adversely affected, and consumer purchasing patterns can change. The COVID-19 pandemic had an adverse effect on the profitability of some businesses. The impact of severe climate-related events in certain parts of the world could be similar.

There are other factors that banks and their corporate clients need to consider. Climate changes are likely to lead to mass migrations as individuals seek to leave

¹⁴In 2019, Pacific, Gas & Electric (PG&E) a California company, filed for bankruptcy because it was blamed for wildfires and lawsuits resulted. This has been dubbed the first climate-related bankruptcy.

unbearably hot climates, areas that have been flooded, areas where clean water is no longer available, or areas that have become unsuitable for farming. Migration can happen within countries or between countries. This can be expected to lead to widespread economic disruption and create difficult problems for governments to handle. The impact on agriculture will be mixed. In some northerly regions, crop yields can be expected to increase as a result of warmer climates. In other areas, crops that have been grown for centuries will no longer be viable because the land is too dry. Rearing healthy animals for food will be more difficult in many parts of the world. Food shortages are likely. Pests that threaten crops grown in different regions will also migrate with the changing climate, and this can cause crop failures and periods of low yields as farmers learn to deal with these problems. Farmers who are desperate for income might require increasing amounts of fertilizer to deal with difficulties caused by climate change. This might accentuate difficulties, as fertilizer production is a major cause of greenhouse gas emission.

Sovereign debt is likely to be affected by climate change. As companies are adversely affected by climate risks, they become less profitable. As a result, tax revenue declines and the demand for borrowing by countries rises. Unfortunately, less-developed countries will tend to be most affected by climate changes and climate-related disasters. Furthermore, these are also countries that need capital to enable them to move away from high-emitting sources of energy (coal and wood burning) and high-emitting agricultural practices. This will present a challenge for banks and the governments of developed countries and may lead to an increase in global tensions.

The transition to a low-carbon economy means that there will be winners and losers. Electric vehicle producers can be expected to do well, as can companies that are concerned with methods of producing energy without fossil fuels (e.g., by using wind or the sun). Oil companies and coal companies can expect difficult times ahead. Royal Dutch Shell was sued by *Millieudefensie* (Friends of the Earth Netherlands), an environmentalist group, and ordered by a court in the Hague in 2021 to reduce its carbon emissions to 45% of 2019 levels by 2030. (Shell is appealing the ruling.) Such lawsuits may become more common. Some observers consider carbon emitters to be analogous to tobacco companies, which have in recent years been on the losing side of lawsuits concerning the impact of cigarettes on a smoker's health.

Banks are concerned about reputational risk. Suppose that a bank considers a loan to a carbon emitter to have low credit risk. It might still refuse to make the loan because it could lead to adverse publicity and environmentally conscious customers closing their accounts. Indeed, banks are coming under pressure from many different organizations to take climate risk into account in their lending decisions. Goldman Sachs, for example, takes reputational risk seriously. In its TCFD report, it states that involvement in industries associated with climate change may lead to “reduced client and employee loyalty, investor divestment and impacts to client activity.” Business Snapshot 23.2 provides another illustration of a financial institution wanting to avoid reputational risks.

BUSINESS SNAPSHOT 23.2**Reputational Risk and HSBC**

An interesting story concerns a *Financial Times* event in May 2022. Stuart Kirk, Global Head of Responsible Investments at the bank HSBC, delivered a speech titled “Why investors need not worry about climate risk.” He first argued that human beings have proved throughout the ages that they are adaptable and manage changes well. He then argued that if an average growth rate in the economy of 2% or more continues until the year 2100, the negative effects of climate change will be manageable because they will be small compared with the gains experienced from economic growth.

Whether you agree with him or not, Mr. Kirk is entitled to his opinion. (News reports after the presentation suggested that many asset managers, and some politicians, would agree with him.) However, HSBC quickly distanced itself from the presentation and suspended him. (This is in spite of the fact that it approved the title of the presentation in advance.) Why did HSBC do this? The reason concerns reputational risk. Financial institutions want to be (and be seen to be) environmentally responsible. It is not in their interest to incur the displeasure of environmental groups and lose business as a result.

In considering arguments such as those of Mr. Kirk, we should keep in mind that climate risks are not like other risks to which the human race has adapted. They are risks that affect the whole planet and its ability to support human life as we know it. It is encouraging that many banks, and their regulators, are trying to act responsibly. The speed and scale of the human response that is required to address climate change means that financial markets and their participants will play an important role in reducing emissions.

23.1.3 Institutional Investors

Institutional investors are increasingly taking climate risk into account in forming portfolios. The \$240 billion California State Teachers’ Retirement System was a pioneer. It formed a Green Initiative Task Force in 2007 to identify, analyze, and propose investment opportunities addressing climate change. The company is active in promoting environmentally friendly behavior (sometimes referred to as corporate social responsibility, CSR) at shareholder meetings.

Companies have an incentive to demonstrate that they are reducing emissions in line with targets. If institutional investors refuse to invest in a certain company because its climate-related disclosures are unsatisfactory, the company is likely to suffer. To put this another way, investors have a great deal of leverage over the companies they invest in and can play a key role in persuading them to reduce greenhouse gas emissions.

Green bonds are an interesting innovation. In 2007, the European Investment Bank issued a “climate awareness bond,” and in 2008 the World Bank used the term *green bond* to describe bonds where the proceeds are used for climate-friendly investments. Green bonds are now issued by both the public and private sector. Investment banks handle the underwriting and an opinion on the environmental credentials of the issuer is often given by an external agency. Other related innovations are green loans and sustainability-linked bonds and loans.

We have already mentioned TCFD, the Task Force on Climate-Related Disclosures, which was formed in 2015 by the Financial Stability Board to develop recommendations on climate-related disclosures. Its recommendations involve four themes: governance, strategy, risk management, and metrics. It recommends that companies disclose:

- How climate risks and opportunities are managed
- The current and potential future impact of climate risks
- The process used to identify the risks
- The metrics used to quantify greenhouse gas emissions

The recommendations have received widespread support. As of October 6, 2021, the Task Force had over 2,600 supporters globally, including 1,069 financial institutions responsible for \$194 trillion of assets. Jurisdictions such as Brazil, the European Union, Hong Kong, Japan, New Zealand, Singapore, Switzerland, and the United Kingdom have introduced or announced their intention to introduce mandatory disclosure requirements that are aligned with TCFD.¹⁵

One issue is that a number of different organizations are trying to increase transparency and measurability of climate-related financial reporting. Similar to the financial reporting industry, coordination is vital. For financial reports, large umbrella organizations like IFRS and FASB led the effort toward global coordination of accounting standards to ensure that investors can understand and rely on corporate reports and avoid falling prey to financial reports that have misleading or fraudulent content. Similarly, organizations like TCFD are working very hard to get sustainability-related reports to be credible and comparable. However, these efforts are still a work in progress, and the attendant uncertainty presents yet another risk to investors.

In June 2021, 457 investors with over \$41 trillion in assets under management signed the 2021 Global Investor Statement to Governments on the Climate Crisis. Others have signed since. The Statement was a call for governments to undertake five actions:

1. Strengthen nationally determined policies to limit warming to 1.5 degrees Celsius.
2. Commit to zero emissions by 2050.
3. Strengthen carbon pricing, phase out fossil fuel subsidies and coal-based power, and avoid new carbon-intensive infrastructure.

¹⁵See The Financial Stability Board, Task Force on Climate-Related Disclosures, Bank for International Settlements, October 2021, https://assets.bbhub.io/company/sites/60/2021/07/2021-TCFD-Status_Report.pdf.

4. Ensure that COVID-19 economic recovery plans are consistent with the transition to net zero emissions.
5. Commit to introducing mandatory climate risk disclosures.

Mark Carney, former governor of the Bank of Canada and the Bank of England, has led an effort to ensure that financial institutions account for and disclose the climate risks in their lending and investments. As United Nations special envoy for Climate Action and Finance, he chaired the Glasgow Financial Alliance for Net Zero (GFANZ), which was formed ahead of the 2021 Glasgow summit. This is a global coalition of leading financial institutions committed to accelerating the decarbonization of the economy.

It should be clear from this that there are many initiatives aimed at using investors to persuade companies to reduce, and eventually eliminate, greenhouse gas emissions. Some asset managers such as BlackRock have embraced the net zero objective enthusiastically. Others such as Berkshire Hathaway, a conglomerate headed by Warren Buffett, have been reluctant to disclose climate risk metrics, but are coming under increasing pressure to do so. Unfortunately, the pressure is not all one way. In July 2022, Florida's governor proposed legislation that would ban the state's pension funds from selecting investments based on environmental factors. In August 2022, attorney generals from 19 U.S. states wrote to BlackRock CEO Larry Fink, asking him to focus on profits and not take environmental issues into account in state pension investments.

To summarize: companies are under pressure (for financial, societal, and legal reasons) to disclose their current emissions and develop plans for reducing them. As just mentioned, not all politicians have been supportive of this. But, in spite of that, there is an increasing focus on emissions, and it is to be hoped that this will result in climate change proving to be manageable.

23.2 ESG

ESG stands for environmental, social, and governance. Whereas it used to be considered that it was the job of a company to maximize the profits for shareholders, it is now recognized that the profit objective must be combined with environmental and social responsibility.¹⁶

ESG is often considered to be synonymous with climate risk, but ESG has somewhat broader concerns than climate. The “S” covers social issues such as diversity, equity, inclusion, and worker well-being. The metrics to measure social factors are not as well developed as those for environmental factors, but it is likely that companies will be required to disclose more information in this area in the future. How does a company manage discrimination and harassment in the workplace? What steps is it taking to encourage gender, racial, and ethnic diversity in its management of human capital?

¹⁶A famous (now dated) quote from Nobel Prize-winning economist, Milton Friedman in 1970 is “There is only one social responsibility of business: to use its resources and engage in activities designed to increase its profits.”

Some large companies in the United States are recognizing that the social landscape is changing and are changing their internal practices. The Securities and Exchange Commission (SEC) has approved a proposal by the stock exchange Nasdaq to require listed companies to have one director who is female and another who is black, Hispanic, Native American, LGBTQ+, or part of another underrepresented group (or to explain why they do not have this level of diversity). In 2020, Goldman Sachs announced that it would stop underwriting IPOs for companies in the United States and Europe that do not have diverse boards.

The “G” in ESG is concerned with companies using accurate and transparent accounting methods and having governance mechanisms in place to ensure policies and practices are actually followed. A company should be accountable to shareholders and avoid conflicts of interest in their choice of board members and senior executives. It should not make political contributions in order to obtain preferential treatment. It should also not engage in illegal activities such as bribing foreign officials to obtain contracts. The CEO and chairperson of the board should be different individuals. Measures of good governance have existed for some time. For example, Daniel Kaufmann from the Brookings Institute and Aart Kray from the World Bank have produced Worldwide Governance Indicators (WGI) for many years.

One of the difficulties in this area, similarly to the area of climate accounting, is that there are many different providers of ESG data. Not surprisingly, these ESG reports from different data providers often disagree with each other. Since there is currently no regulation concerning ESG reporting standards, and since ESG data providers are competing for business by creating reports that users will pay for, there are many ways for firms to achieve a good ESG rating. Firms can, effectively, shop around for a good ESG score. This lack of standardization and transparency represents a source of risk for investors.

23.3 Sustainability

Like ESG, sustainability includes climate risk, but covers other issues. In December 1983, the secretary-general of the United Nations, Javier Pérez de Cuéller, asked the former prime minister of Norway, Gro Harlem Brundtland, to create an organization that became known as the Brundtland Commission. Its task was to consider environmental strategies for sustainable development. The Brundtland report’s definition of sustainable development has been widely used:

Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their needs.

The production of greenhouse gases is an important example of nonsustainable development, but there are others. The way we dispose of waste is an important consideration. The pollution of oceans with plastic waste, manufactured chemicals, petroleum

waste, agricultural runoff, sewage, and more is a particular concern. It kills fish and exposes humans to toxins because they eat contaminated seafood. Another example of unsustainable development is the contamination of soil and water sources from pesticides, fertilizers, petroleum products, and so on.

The Sustainability Accounting Standards Board (SASB) was founded in 2011 as an independent nonprofit organization to provide sustainability metrics. Another similar initiative is the Global Reporting Initiative, which was founded in 1997 following the Exxon *Valdez* oil spill.

There are considered to be three pillars of sustainability:

- Environmental sustainability
- Economic sustainability
- Social sustainability

We have discussed environmental sustainability. Economic sustainability refers to practices that support the long-term development of a company or nation while also protecting environmental, social, and cultural elements. Social sustainability occurs when today's actions support the capacity of current and future generations to create healthy and livable communities. Sustainability requires that we consume natural resources at a rate where they are able to replenish themselves. Overfishing or deforestation are examples of activities that are not sustainable.

A number of funds with the explicit objective of investing in companies with good ESG or sustainability practices have been developed. They are designed to appeal to retail investors who want to use their investment dollars to make the world a better place. An interesting question is: Do ESG/sustainable funds outperform other funds? Although the demand for these funds has increased, their performance appears to be no better than that of other funds.¹⁷ Unfortunately, some organizations produce reports with questionable research practices that try to convince their customers that “doing well by doing good” is easy and will lead to high returns with low risks. Finance theory suggests that high returns tend to be accompanied by high risks. In any event, earning high returns should not be a primary goal for investing in sustainability-motivated funds.

23.4 Greenwashing

Companies in many industries are keen to present themselves as ecologically responsible. Consider two similar products, one advertised as environmentally friendly and the other promoted in a different way. Many consumers would prefer to buy the first product. This has led to some misleading claims known as *greenwashing*. Here are some

¹⁷See S. M. Hartzmark and A. B. Sussman, “Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows,” *Journal of Finance* 74 (6) (December 2019): 2789–2837.

examples of greenwashing from Akepa, an organization that works with companies to create sustainable and ethical brands:¹⁸

- Innocent drinks, owned by Coca-Cola, suggested in its TV advertising that its plastic products were environmentally friendly when in fact they could not be recycled.
- Keurig made misleading recycling claims in its packaging about plastic pods. It was fined \$3 million and ordered to change what it says.
- IKEA has generally good sustainability credentials, but ran into problems because its beechwood chairs were using illegally sourced wood.
- Ryanair claimed, with no evidence, that it was Europe's "lowest emissions airline."

Unfortunately, financial institutions do not have a perfect record as far as greenwashing is concerned. In 2022, Deutsche Bank's asset management subsidiary, DWS, was found guilty of exaggerating the green credentials of the investments it sold. To quote Magdalena Senn of the German consumer advocacy group Finanzwende, "Greenwashing is not a trivial offense." The CEO of DWS was forced to resign.

Summary

Risk management groups now actively consider environmental risks. At most large organizations, there is a senior executive responsible for these risks. As the extent of the risks to our planet become increasingly apparent, environmental risks are likely to become an increasingly important consideration for many companies and countries. The Global Association of Risk Professionals (GARP) now offers a Sustainability and Climate Risk (SCR) qualification in addition to its well-known FRM. The CFA Institute offers a Certificate in ESG Investing in addition to the CFA charter.

Further Reading

Garbe, J., T. Albrecht, A. Levermann, J. F. Donges, and R. Winkelmann, "The hysteresis of the Antarctic Ice Sheet." *Nature* 585 (September 2020): 538–544.

Gates, W. *How to Avoid a Climate Disaster: The Solutions We Have and the Breakthroughs We Need*. Random House, 2021.

Golnaraghi, Maryam. *Climate Change Risk Assessment for the Insurance Industry. The Geneva Association*, 2021. <https://www.genevaassociation.org/research-topics/climate-change-and-emerging-environmental-topics/climate-change-risk-assessment>.

Hartzmark, S. M., and A. B. Sussman. "Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows." *Journal of Finance* 74 (6) (December 2019): 2789–2837.

IPCC, Sixth Assessment Report, 2021. <https://www.ipcc.ch/assessment-report/ar6/>.

OECD. *Financial Markets and Climate Transition: Opportunities, Challenges and Policy Implications*. Paris: OECD. <https://www.oecd.org/finance/Financial-Markets-and-Climate-Transition-Opportunities-challenges-and-policy-implications.htm>. 1990.

¹⁸See "Greenwashing: 10 Recent Stand-Out Examples," Akepa, July 23, 2021, <https://thesustainableagency.com/blog/greenwashing-examples>.

Practice Questions and Problems (Answers at End of Book)

- 23.1 Why does the burning of fossil fuels lead to temperature rises?
- 23.2 What is the goal that was agreed to in Paris in 2015?
- 23.3 What are SSPs?
- 23.4 What is the difference between a carbon tax and a cap and trade system?
- 23.5 What is the difference between physical risks and transition risks?
- 23.6 What are the risks of climate change to insurance companies?
- 23.7 What does ESG stand for?
- 23.8 What is greenwashing?

Further Questions

- 23.9 How can pension plans and mutual funds help solve the global warming problem?
- 23.10 Give examples of activities covered by ESG and sustainability that are not climate risk issues.

Chapter 24

Enterprise Risk Management

Much of this book has focused on understanding and quantifying different types of risk: market risk, credit risk, operational risk, and so on. An important part of a risk manager's job is to ensure that these risks are correctly evaluated. But it is also important for the risk management function to take a holistic, big-picture view of risk. It should identify potential adverse events and their full consequences. The total exposure to an adverse event can be greater than (or less than) the result obtained by considering each risk type separately. Enterprise risk management (ERM) is the name given to this holistic approach to risk management.

In understanding ERM, it is important to distinguish top-down and bottom-up approaches to risk management. Bottom-up approaches are concerned with assessing the different types of risk borne by different business units and combining them. In top-down approaches, the overall risk appetite of the organization is defined, and this is then used to define risk limits for different parts of the organization. In practice, a financial institution needs to use both top-down and bottom-up approaches. A top-down approach is necessary to define the overall risk appetite, and a bottom-up approach is necessary to evaluate whether the risks being taken by business units are consistent with this risk appetite.

In 2004, the Committee of Sponsoring Organizations of the Treadway Commission (COSO) issued the Enterprise Risk Management Integrated Framework (see www.coso.org) and proposed this definition of ERM:

Enterprise risk management is a process, effected by an entity's board of directors, management, and other personnel, applied in strategy setting and across the enterprise, designed

to identify potential events that may affect the entity, and manage risk to be within its risk appetite, to provide reasonable assurance regarding the achievement of entity objectives.

This definition makes a number of points. First, there should be board involvement in ERM. Second, ERM is a component of a company's strategy. Third, it involves identifying potential adverse events. Fourth, as part of ERM the company must identify its risk appetite and manage risks in a way that is consistent with that appetite. Fifth, ERM should help a company achieve its objectives and be a core part of strategic planning and strategic execution processes.

Many financial institutions (and their boards) have historically had a compliance orientation as far as top-down risk management is concerned. Their aim was to satisfy regulators and stay within the letter of the law. ERM is moving financial institutions away from this mind-set. It is leading to the risk management function having a strategic orientation and being value enhancing. A company that manages risk effectively has a competitive advantage over another company that does not do so.

This chapter considers the essential elements of ERM. It considers risk appetite, risk culture, how potential adverse events are identified, and how ERM is integrated into a financial institution's strategy.

24.1 Risk Appetite

An essential part of ERM is concerned with specifying the organization's risk appetite. How much risk is the organization prepared to take in order to achieve its strategic objectives and business plans? One dimension of the risk appetite of a financial institution is likely to be concerned with the loss it is prepared to incur if a worst-case scenario happens. This can be expressed in terms of an enterprise-wide value at risk (VaR) or expected shortfall (ES) measure.

For a fund manager, this dimension of risk appetite is relatively easy to quantify—at least in theory. Chapter 1 explained the theoretical trade-off between expected return and risk. The mix of assets influences the beta of the portfolio, and this in turn determines the theoretical expected return and risk.¹ Figure 24.1, which is similar to Figure 1.4 in Chapter 1, shows the trade-off between risk and expected return for a portfolio. As shown in Section 1.2, if the beta of the portfolio is β , the expected return μ_P is

$$\mu_P = (1 - \beta)R_F + \beta R_M$$

and the standard deviation of the return is

$$\sigma_P = \beta \sigma_M$$

¹In practice, as explained in Chapter 1, the theoretical return for a particular value of beta is often used as a benchmark. Fund managers who beat the benchmark are adding value.

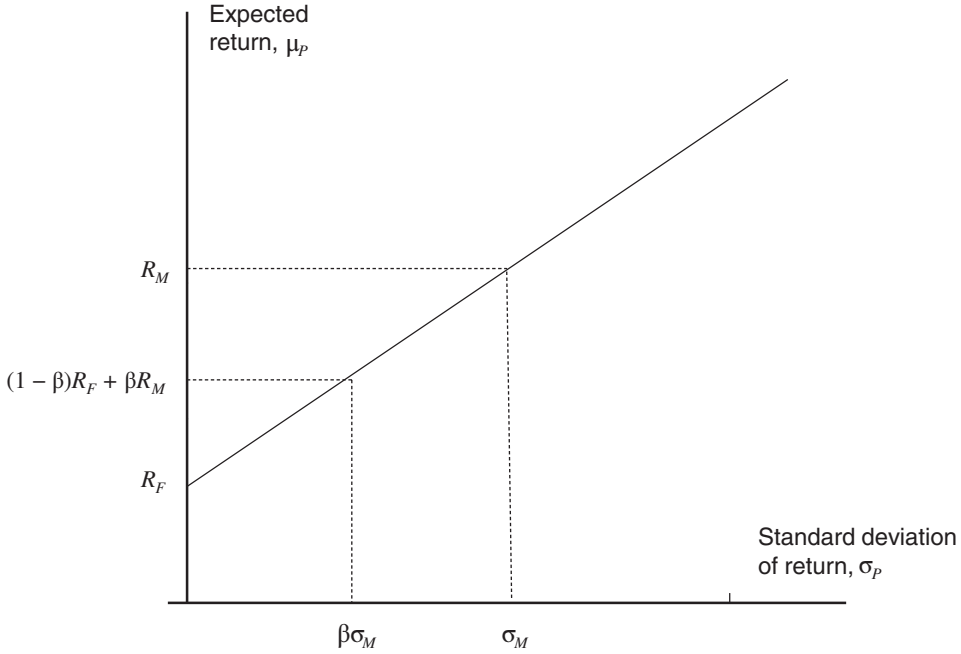


Figure 24.1 Dependence of Mean and Standard Deviation of a Portfolio on Beta, β

where σ_M is the standard deviation of the market return, R_F is the risk-free return, and R_M is the expected return on the market. Assume returns are normally distributed.² The return, R , that has a probability p of being exceeded in one year is given by

$$N\left(\frac{R - \mu_p}{\sigma_p}\right) = 1 - p$$

where N is the cumulative normal distribution function. Substituting for μ_p and σ_p we obtain

$$\frac{R - (1 - \beta)R_F - \beta R_M}{\beta \sigma_M} = N^{-1}(1 - p)$$

or

$$\beta = \frac{R - R_F}{R_M - R_F + N^{-1}(1 - p)\sigma_M} \quad (24.1)$$

This formula allows us to determine the beta that is consistent with a particular VaR objective. Table 24.1 shows returns including dividends from the S&P 500 over the years between 1970 and 2021. The mean and standard deviation of returns over the period are 12.45% and 16.88% so that $R_M = 0.1245$ and $\sigma_M = 0.1688$. Suppose that the risk-free rate is 2% ($R_F = 0.02$). Suppose further that a fund's risk appetite is such that it wants to

²This assumption is made to provide a simple illustration. The returns from the market have heavier tails than the normal distribution, so in practice the returns would be fitted to a different distribution, perhaps a Student's t -distribution with a relatively small number of degrees of freedom.

Table 24.1 Total Annual Return from the S&P 500 (including dividends) between 1970 and 2021

Year	Return (%)	Year	Return (%)	Year	Return (%)
1970	4.01	1988	16.61	2006	15.79
1971	14.31	1989	31.69	2007	5.49
1972	18.98	1990	−3.10	2008	−37.00
1973	−14.66	1991	30.47	2009	26.46
1974	−26.47	1992	7.62	2010	15.06
1975	37.20	1993	10.08	2011	2.11
1976	23.84	1994	1.32	2012	16.00
1977	−7.18	1995	37.58	2013	32.39
1978	6.56	1996	22.96	2014	13.69
1979	18.44	1997	33.36	2015	1.38
1980	32.50	1998	28.58	2016	11.96
1981	−4.92	1999	21.04	2017	21.83
1982	21.55	2000	−9.10	2018	−4.38
1983	22.56	2001	−11.89	2019	31.46
1984	6.27	2002	−22.10	2020	18.40
1985	31.73	2003	28.68	2021	28.71
1986	18.67	2004	10.88		
1987	5.25	2005	4.91		

have a 95% probability of not losing more than 10% in one year. Substituting $R = -0.1$ and $p = 0.95$ into equation (24.1), the fund manager's beta should be

$$\beta = \frac{-0.1 - 0.02}{0.1245 - 0.02 + N^{-1}(0.05) \times 0.1688} = 0.69$$

This means that to achieve the VaR objective the risk manager should invest 69% of the portfolio in the market and the rest in a risk-free asset.³

Unfortunately, the available risk-return alternatives for a financial institution such as a bank or an insurance company cannot usually be quantified as easily as for a fund manager. When markets are efficient, we can use the arguments in Chapter 1 to show that for stock market and similar investments an efficient frontier similar to that in Figure 24.1 should exist. But there is no reason why this efficient frontier should define trade-offs between risk and expected return for major strategic investments undertaken by a bank, insurance company, or other type of financial institution. Some strategic investments will be above the efficient frontier and some will be well below.

For any company, strategic investments that are above the efficient frontier are usually those that take advantage of the company's competitive advantages. Companies, whether financial or non-financial, should therefore define their competitive advantages and look

³In practice, a fund manager is likely to want to incorporate his or her judgment about returns into the choice of a portfolio. A way this can be done is provided by F. Black and R. Litterman, "Asset Allocation: Combining Investor Views with Market Equilibrium," *Journal of Fixed Income* 1, no. 2 (September 1991): 7–18.

for strategic investments that exploit those advantages. There is some evidence that they have not always been very successful at doing this. A well-known study by Bowman uncovered a negative relationship between risk and return for companies in most sectors.⁴ This is sometimes termed the *Bowman paradox*. In a further study, Bowman showed that more troubled firms often take less justifiable risks.⁵ One explanation for this might be that a company experiencing difficulties often has few competitive advantages. It may be tempted to undertake a major project that has only a small probability of success if success would ensure survival.

The market in which financial institutions operate is highly competitive, and competitive advantages are not as common as they are in some other industries. Typically, attempts by a financial institution to increase returns involve an increase in systematic (beta) risk, and the financial institution's shareholders can often achieve an equivalent (and sometimes better) risk-return trade-off for themselves. The main competitive advantage of a financial institution is the trust of its customer base (retail and commercial) that it has built up. Providing services to that customer base should give a satisfactory risk-return trade-off. For regulated financial institutions, the risk-return trade-off is impacted by increasingly stringent capital and liquidity requirements.

As mentioned, the part of a financial institution's risk appetite concerned with the worst-case loss, or equivalently the worst-case erosion of equity capital, can be expressed by specifying VaR or expected shortfall with a particular time horizon and confidence level. Other parts of the risk appetite are less easy to quantify. They are likely to be concerned with reputational risk, credit rating risk, the risk of not complying with regulations, legal risks, and so on. Some of these are likely to be expressed qualitatively rather than quantitatively. For example, a bank might state that any project that could appreciably reduce its reputation among its core customers, and how trustworthy it is perceived to be by those customers, is unacceptable.

Business Snapshot 24.1 describes how Wells Fargo put employees under a huge amount of pressure to meet unrealistic sales targets in order to increase sales. This strategy led to employees engaging in questionable and sometimes fraudulent activities in order to meet the targets. The strategy should have been rejected (or at least monitored more carefully) because of its potential adverse effect on how customers perceived the bank and because of its adverse effect on the bank's relationship with its employees.

The quantification of risk has been a major theme running through this book. Up to now we have implicitly tended to take a bottom-up approach. We have looked at different ways risk measures can be calculated by business units and then considered how the risk measures can be combined to produce a risk measure for the whole enterprise. The specification of a risk appetite framework is a top-down activity that should involve

⁴Bowman, E. H., "A Risk-Return Paradox for Strategic Management," *Sloan Management Review* 21 (1980): 17–31.

⁵Bowman, E. H., "Risk Seeking by Troubled Firms," *Sloan Management Review* 23 (1982): 33–42.

BUSINESS SNAPSHOT 24.1

Cross-Selling at Wells Fargo

Wells Fargo is a bank that survived the 2007–2008 crisis well. It avoided the risks taken by other banks during the period leading up to the crisis and gained a reputation for stability. Unfortunately, subsequent events tarnished its reputation and led to a loss of trust from potential clients.

Cross-selling involves attempting to sell multiple products to clients. A client with a checking account might be encouraged to take out a credit card or open a savings account. Wells Fargo was considered to be one of the best banks at cross-selling, and indeed there is nothing wrong with cross-selling. But if it is carried out too aggressively it can be bad for a bank's reputation and even be fraudulent.

Wells Fargo considered branch employees to be “salespeople” and clients to be “customers.” Branch personnel were given very aggressive quotas for the number of products they were expected to sell to clients, and they were under a huge amount of pressure. As a result, they ordered credit cards for clients without the clients' consent. They also created checking and savings accounts for clients who had not asked for them. In many cases, these activities led to clients paying additional fees to the bank. Quotas were sometimes even satisfied by enrolling homeless people in fee-accruing products, even though they would almost certainly fail to make payments as they became due.

Wells Fargo's dysfunctional sales culture was first documented by the *Wall Street Journal* in 2011 and further described in an article in the *Los Angeles Times* in 2013. Some reforms were then made, but the bank was fined \$185 million in September 2016 for creating more than 1.5 million unauthorized deposit accounts and over 0.5 million unauthorized credit-card accounts between 2011 and 2016. Later estimates in May 2017 suggested that the number of unauthorized accounts was higher than this, closer to 3.5 million in total.

Senior executives claimed that they were unaware of what was going on as a result of the company's pressure-cooker sales culture. This seems unlikely given that employees made calls to the company's ethics hotline and complaints from customers had to be dealt with. The bank fired 5,300 employees between 2011 and 2016 as a result of the fraudulent sales practices. The CEO, John Stumpf, agreed to forgo \$41 million in stock options and resigned in October 2016. Later, \$28 million of his earnings were clawed back. Other executives also lost earnings. The bank paid \$110 million to clients who had accounts opened without their permission. Investigations were commenced by the Senate Banking Committee, the Securities and Exchange Commission, and prosecutors in a number of jurisdictions.

Clearly, the aggressive sales culture at Wells Fargo was misguided. It alienated employees, many of whom were under unreasonable stress and found it difficult to find new jobs when they were fired. It also destroyed the trust and good reputation that Wells Fargo had built up with its clients over many years. If Wells Fargo's risk appetite had included a desire to avoid doing activities that could adversely affect its reputation, it would surely have reduced the sales pressure and emphasized the importance of developing good long-term relationships with clients.

Wells Fargo took steps to change its sales culture under a new CEO in late 2016. It stopped the practice of giving employees sales quotas. Some employees who had been fired were rehired. It has taken time for the bank's relationships with employees and clients to recover, but fortunately Wells Fargo is financially strong enough to bounce back.

the board of directors. Risk appetite is first defined for the institution as a whole. It is then used to define risk limits for business lines. This can involve some experimentation where bottom-up calculations are used to see if the risk appetite of the organization is being met. For example, a financial institution wanting to preserve its AA rating might decide that the probability of losses exceeding \$50 billion in one year must be 0.03% or less. It then needs to determine by trial and error what the VaR limits for its trading activities, its lending activities, and so on must be so that when aggregated they are consistent with this aspect of the risk appetite. Once risk limits have been set, it is important to monitor the decisions taken in the business units to ensure that the limits are being adhered to.

One important element of risk appetite is risk concentration. With the benefit of hindsight we can say that some large banks had portfolios that were too concentrated in subprime mortgages in 2006 and 2007. The risk appetite statement should directly or indirectly have implications for concentration risk. It should limit credit exposure to any one company, limit credit exposure to any one sector of the economy, limit credit exposure to movements in any one market variable, and so on.

Not surprisingly, regulators are interested in encouraging banks to develop risk appetite frameworks. Indeed, bank supervisors are required to ensure that effective risk appetite statements have been developed and implemented. In 2013, the Financial Stability Board published a document that contained definitions of relevant terminology and the principles on which a risk appetite framework should be based.⁶ The document

⁶See Financial Stability Board, "FSB Principles for an Effective Risk Appetite Framework," Consultative Document, July 2013.

spells out the responsibilities of the board of directors, the chief executive officer, the chief risk officer, the chief financial officer, and business line managers.

24.2 Risk Culture

A financial institution's risk culture is all about how decisions are made. If the risk culture is good, all decisions are made in a disciplined way with a careful consideration of possible outcomes, weighing risks and rewards. It should be emphasized that this does not mean that no risks are taken. It means that risks are considered in relation to potential rewards and the risk appetite of the organization. The Financial Stability Board has suggested the way bank supervisors can assess risk culture at large systemically important banks.⁷

Decisions often involve trade-offs between short-term profits and medium- to long-term risks. A decision that is certain to lead to an immediate profit sometimes has the disadvantage that it is liable to give rise to serious problems later. Consider the situation of a salesman who works for a financial institution and is deciding whether to sell a sophisticated financial product to a customer when he knows it does not meet the customer's true needs. The sale would be profitable to the financial institution and would increase the year-end bonuses of the salesman and the group that developed the product. But it might have adverse future consequences. For example, the reputation of the financial institution could suffer if the customer incurs losses as a result of the product. Expensive lawsuits might follow.

The product described in Business Snapshot 5.4, which was sold by Bankers Trust to Procter & Gamble, is a classic example of what can go wrong. The product had a fairly high probability of saving Procter & Gamble 75 basis points on its funding and a small probability of costing the company a huge amount of money. The scenario where there is a huge cost to Procter & Gamble may or may not have been fully explained by the Bankers Trust salesman. But, if there had been a good risk culture within Bankers Trust, management would have questioned whether it made any sense for it to sell the product at all. Indeed, the quants working for Bankers Trust would have been told not to develop the product in the first place.

Procter & Gamble was one of many companies that entered into transactions similar to the one described in Business Snapshot 5.4 with Bankers Trust. As it happens, the low-probability outcome materialized and the transactions proved to be very costly for Bankers Trust's clients. There were numerous lawsuits, all of which were settled out of court. Bankers Trust, a company that had built an enviable reputation in financial markets, was ruined and later acquired by Deutsche Bank. Of course things could have worked out well and Bankers Trust's clients could have been grateful to the bank for saving them 75 basis points on their funding. But the risk to Bankers Trust and its reputation of huge client losses was clearly too great. The risk culture was one where employees thought only about short-term profits and not about the longer-term repercussions.

⁷See Financial Stability Board, "Guidance on Supervisory Interaction with Financial Institutions on Risk Culture: A Framework for Assessing Risk Culture," April 2014.

BUSINESS SNAPSHOT 24.2

The Santander Rail Deal

In 2007, the Portuguese rail company Metro do Porto (MdP) was looking for a way to reduce its 4.76% per annum funding costs. After consulting a number of banks, it agreed to a 14-year swap with the Spanish bank Santander. In this swap, Santander paid the 4.76% funding cost and in return MdP paid 1.76% plus a spread. For the first two years, the spread was zero, saving MdP 3% on its funding. After that, the spread was calculated each quarter as

$$\max[0\%, \text{Previous Spread} + 2 \times \max(2\% - R, 0) + 2 \times \max(R - 6\%, 0) - D]$$

where R is the three-month Euribor rate and D , referred to as the DigiCoupon, is 0.5% if R lies between 2% and 6%, and zero otherwise.

At the start of the deal, three-month Euribor was about 4%. If it had stayed between 2% and 6% for the whole 14 years, MdP would have saved 3% per annum on its funding for the whole period. As it happened, three-month Euribor had dropped below 2% by 2009 when the deal was only two years old, and it stayed below 2% for several years. As a result, the spread increased rapidly, reaching 40.6% in September 2013. The problem for MdP was that the spread in one quarter is dependent on the spread in the previous quarter. Suppose that three-month Euribor is 0.5% from year 2 onward. MdP's spread would increase by 3% per quarter or 12% per year. If subsequently three-month Euribor increases so that it is in the 2% to 6% range, the spread declines, but only at the rate of 0.5% per quarter or 2% per year.

The deal was subject to a lawsuit. (Disputes over other similar deals entered into by Santander with other Portuguese state-owned entities were settled out of court.) MdP should of course have run scenario analyses to understand the potential costs of the transaction. Whether it did this and whether it understood the cumulative nature of the spread adjustments is not known. Possibly, it overweighted the attractiveness of saving 3% per year for the first two years. According to an article in *Risk* magazine in 2014, one London-based corporate hedging adviser has described this deal as a contender for the worst trade of all time!

Another example is provided by the transaction between the Spanish bank Santander and the Portuguese state-backed rail operator Metro do Porto in 2007, described in Business Snapshot 24.2. The transaction allowed the rail operator to reduce its funding costs for the first two years, but this was at the risk of very high funding costs for later years. From Metro do Porto's perspective, the trade-off being made between short-term

BUSINESS SNAPSHOT 24.3

ABACUS

In 2007, Goldman Sachs created a product called ABACUS from a synthetic ABS CDO. As explained in Chapter 6, an ABS CDO is formed from a portfolio of tranches (often mezzanine tranches) of ABSs. The word *synthetic* means that Goldman defined the ABS CDO, but it did not actually buy the underlying tranches. It then acted as an intermediary between two parties. Party A bought a credit default swap providing protection against what the losses on the senior tranche of the ABS CDO would have been if it had been created. Party B sold the credit default swap, agreeing to provide the protection. Party A paid Party B an insurance premium (the CDS spread).

So far, so good. This is what investment banks such as Goldman Sachs do all the time. They act as intermediaries between buyers and sellers. In this particular case, Party A was a hedge fund controlled by John Paulson. This was one of the few hedge funds to make big bets against the performance of subprime mortgages and the housing market generally. It transpired that Paulson paid Goldman Sachs \$15 million for creating the structure and that his hedge fund had a hand in choosing the ABSs that went into the structure. Party B was IKB Deutsche Industriebank AG (IKB), a bank based in Dusseldorf, Germany, that specializes in lending to small and medium-size companies. The portfolio manager was ACA, which at the time had a good reputation for managing ABS CDOs.

It is alleged that Goldman Sachs represented to IKB that the ABS CDO had been defined by ACA when the ABSs in the portfolio had in fact been chosen by John Paulson's hedge fund as ones that were likely to incur huge losses. Specifically, it is alleged that Paulson constructed the synthetic ABS CDO with tranches formed from portfolios of mortgages where the borrowers had low FICO scores and came from states where the rate of house appreciation had been the highest.

Both IKB and ACA took huge losses on ABACUS, and Paulson's hedge fund made a huge profit. (IKB had to be bailed out because of the massive losses it sustained from the subprime meltdown.) The deal was investigated by the Securities and Exchange Commission (SEC), and Goldman Sachs ended up paying a fine of \$550 million, the largest in the history of the SEC at that time. Goldman Sachs's position was hurt when it was revealed that an executive working on the deal, Fabrice Tourre, had sent the following e-mail to a friend: "More and more leverage in the system. The whole building is about to collapse anytime now. . . . Only potential survivor, the fabulous Fab . . . standing in the middle of all these complex, highly leveraged, exotic trades he created without necessarily understanding all the implications of those monstrosities!!!"

gain and long-term risk was highly questionable. But the same was true of Santander. This was one of many similar products sold by Santander to Portuguese state-owned entities. No doubt the products were viewed as profitable by Santander at the time it entered into them, but, as later settlements and lawsuits show, subsequent costs and the negative impact on Santander's reputation have the potential to outweigh the short-term profits.

Yet another example of the conflicts that can exist between short-term profits and long-term risks is provided by Business Snapshot 24.3. This describes a deal, ABACUS, that Goldman Sachs sold to one of its clients without, it is alleged, fully explaining the risks and how it had been created. This is another example of short-term profit being booked without regard for longer-term reputational risk and costs.

The ABACUS deal and one or two other events have led Goldman Sachs to embark on a program to change its corporate culture. Everyone at the top agreed that change was necessary if Goldman was to win back the allegiance of its corporate and investment clients. The CEO, Lloyd Blankfein, traveled to 23 countries during an 18-month period during which he stressed that ethics and reputation were as important as making money (i.e., short-term profits should not be taken if there are potential adverse long-term implications). Gary Cohn was appointed head of the Firmwide Client and Business Standards Committee. One thing this committee did was assess "whether our clients had the background experience and capacity to understand the range of outcomes from transactions they execute with us." The Firmwide Client and Business Standards Committee took steps to ensure that Goldman's bankers and traders were interfacing honestly with clients and avoiding any questionable activities. Consensus is that the reputation of Goldman Sachs has improved as a result of the changes.

In financial institutions, compensation plays a key role in determining the trade-offs between short-term profits and long-term risks. Traditionally, bonuses have been paid once a year to employees based on their performance during the year. This meant that the time horizon for decision making was usually the next bonus date (always less than a year). If a deal could lead to adverse consequences but not before the year end, the possible adverse consequences were likely to be ignored. There has been an attempt in many financial institutions to change compensation arrangements to address this short-term focus. The payment of the bonus for a year is now often spread out over several subsequent years, and bonuses can be clawed back (i.e., not paid) if subsequent events indicate that the year was not as profitable as thought.⁸

24.3 Identifying Major Risks

An important part of ERM is identifying key exposures for both a financial institution's existing business and any proposed strategic investments. The identification of these

⁸Legal issues can make it difficult for financial institutions to try to recover bonuses that have already been paid. Hence, clawback schemes have to be combined with bonus deferral.

exposures is in many ways similar to the choice of scenarios for stress testing. History can provide some guidelines. For example, a financial institution should consider how it would fare if a recession as bad as that experienced during 2007 to 2009 were to occur. The impact of current and emerging trends should also be carefully analyzed. Economists and senior management working for the financial institution can provide useful insights. As indicated in earlier chapters, the environment is changing for banks. One of the results of the credit crisis that started in 2007 is that banks are becoming more tightly regulated. Some activities such as proprietary trading can (at least in the United States) no longer be undertaken. Capital requirements and liquidity requirements are higher. Some financial intermediation business is moving from banks to shadow banks such as money market funds, mortgage companies, securitization vehicles, and so on. In many cases, shadow banks are less heavily regulated than banks and, as a result, can offer more competitive services.

24.3.1 *Cognitive Biases*

Our ability to identify risks effectively is affected by what are termed *cognitive biases*. These describe the tendencies for human beings to think in certain ways and be less than perfectly rational. One common cognitive bias is wishful thinking. It is sometimes difficult to distinguish between what we want to happen (e.g., a project to be a success) and what we think will happen. (Try asking a Manchester United supporter to estimate the chance of Manchester United winning the FA cup next year!) When we want something to happen, we are liable to think only of reasons why it will happen.

Over 100 cognitive biases have been listed by psychologists. Much of the pioneering work was done by Daniel Kahneman and Amos Tversky.⁹ Kahneman won the Nobel prize for economics in 2002 for his work with Tversky on prospect theory, which is concerned with the way people choose between risky alternatives. (Tversky had died a few years earlier.)

One important bias is *anchoring*. When evaluating a potential outcome (e.g., the revenue resulting from a major new venture), we are liable to anchor onto the first estimate that is made. We tend to make relatively small adjustments to that estimate (this is referred to as “anchoring and adjustment”) and often never consider the full range of possible outcomes. In particular, important adverse outcomes may implicitly be considered to have no chance of occurring. To illustrate anchoring, one could ask a group of people to make a best guess of something that is unknown to them such as, perhaps the population of Iceland. They can then be asked to provide a range consisting of the 5th percentile to 95th percentile of their subjective probability distribution. If their estimates are good, the true population of Iceland should lie outside the range only 10% of the time. In practice, it is found that this happens much more frequently. Anchoring causes people to behave as though they know more than they do.

⁹See, for example, D. Kahneman, P. Slovic, and A. Tversky, *Judgment Under Uncertainty: Heuristics and Biases* (New York: Cambridge University Press, 1982).

Another cognitive bias is *availability*. This is where recent information is given undue weight. Sadly, enterprise risk management—and risk management generally—can suffer from availability. Prior to the credit crisis, risk managers in some financial institutions were often not listened to because recent experience had been good. After the crisis, risk managers have had more influence, but as memories of the crisis fade, the “good times will last forever” attitude may return.

Another cognitive bias is known as *representativeness*. This is where an individual categorizes a situation based on a pattern of previous experiences or beliefs about the underlying scenario. It can be useful when trying to make a quick decision, but it can also be limiting because it leads to close-mindedness and stereotyping. Based on previous experience, a senior manager at a financial institution might consider it almost impossible for any other financial institution to successfully compete with it in a particular market. However, if the manager’s past experience is limited, the previous situations might not be representative of future scenarios.

A more subtle mistake sometimes made in estimating probabilities is inverting the conditionality. Suppose that 1 in 10,000 people have a particular disease. A test that is always 99% accurate gives you a positive result (suggesting that you have the disease). What is the chance that you have the disease? Your immediate response is likely to be 99%. However, the true answer is actually about 1%! We are interested in

$$\text{Prob}(D|TP)$$

where D indicates that you have the disease and TP indicates a positive result from the test.

Because the test is always 99% accurate

$$\text{Prob}(TP|D) = 0.99$$

but

$$\text{Prob}(D|TP) \neq \text{Prob}(TP|D)$$

Out of 10,000 people there will be about 100 positive results on average but only one person with the disease.¹⁰ Hence the probability we are interested in is about 0.01. This is an application of the result in probability theory known as Bayes’ theorem.

Yet another bias is the *sunk costs bias*. Suppose that a financial institution has already spent \$1 billion trying to enter a new market. Things are not going well and there seems to be very little prospect of success. Should the \$1 billion influence the financial institution’s decision making? The answer is the \$1 billion is what accountants refer to as a sunk cost. Regardless of the decisions taken now, it cannot be recovered. The key

¹⁰If there are 100 positive results, the test is correct for 9,901 people out of 10,000 so that the accuracy is about 99%.

issue is whether future profits will be sufficiently high to justify future expenditures. In practice, many people are reluctant to admit mistakes, and they continue with projects that are clearly failures for too long. Irrationally, they want to try to get back money already spent, even when the chance of this is very small.

Understanding these biases may assist decision making and the identification of key risks. It should be noted, however, that experiments have shown that it is extremely difficult to eliminate biases. Even when biases such as anchoring are carefully explained and subjects are given financial incentives to make good estimates, the biases persist.

The challenge for ERM is identifying tail risks and trying to estimate the probabilities associated with the adverse scenarios giving rise to the tail risks as well as possible. The cognitive biases we have discussed (and many others have been documented) suggest that the risks will be underestimated. Nassim Taleb makes this point in a best-selling book.¹¹ He is particularly critical of the use of normal distributions for calculating risk measures (as we have been in this book) and argues that extreme events such as the crash of 1987 or the credit crisis of 2007 to 2009 are more likely than many people think.

Using committees of senior management and economists to define adverse scenarios can be useful. Encouraging employees to play the role of devil's advocate when ideas are discussed can also be important. Risks are obviously less likely to be considered if there is an authoritarian CEO who is convinced that he or she is right on major strategic decisions and does not encourage contrary views to be expressed.

24.4 Strategic Risk Management

Once a company's strategy has been defined, strategic risk management involves considering what assumptions have been made and what can go wrong. What will competitors do? How will customers react? How could the strategy be impacted by adverse external events? To quote Michael Porter:¹² "Risk is a function of how poorly a strategy will perform if the 'wrong' scenario occurs."

Once a particular risk has been identified, either in the existing business or in a new strategic investment, there are a number of alternative courses of action:

Avoid: Exit the business/investment.

Reduce: Take action to reduce the probability of the adverse event or its impact.

Adjust: Modify plans to reduce risks.

Share or insure: Transfer or share a portion of the risk.

Accept: Take no action.

Which of these actions is taken is likely to depend on the situation, but it is important that the risk is identified and carefully considered, rather than being ignored.

¹¹See N. N. Taleb, *The Black Swan: The Impact of the Highly Improbable* (New York: Random House, 2007).

¹²See M. E. Porter, *Competitive Advantage* (New York: Free Press, 1985), 476.

Suppose that a successful Canadian bank is considering an expansion of its retail operations by buying a smaller bank in the United States. It has prepared plans indicating that the expansion will increase shareholder value and provide a springboard for increasing its customer base in the United States. There are a number of risks that it needs to consider. Will it be able to transfer its success in Canada to a more competitive banking environment? Will it be able to retain the customers of the bank it is acquiring in the United States? Will it be able to retain key managers? What will happen if there is a severe two-year recession soon after the acquisition is completed? The Canadian bank needs to make sure the acquisition is consistent with its risk appetite. The risk appetite might require it to ensure that if a worst-case scenario happens it will survive without its core Canadian business being appreciably affected. It needs to ensure that the risks can be managed and monitored and that if things work out badly there is a good exit strategy.

Summary

Enterprise risk management is an attempt to take a holistic approach to risk management rather than addressing risk in silos. The central idea is that business units within the financial institution should be part of an integrated, strategic, and enterprise-wide risk management system. The risk appetite for the organization should be defined at the top and approved by the board. Steps should then be taken to ensure that the management of different types of risks in different business units is consistent with the overall risk appetite framework.

A financial institution's risk appetite statement should define how much risk the organization is prepared to take in pursuit of its objectives. Some aspects of the risk appetite, such as the amount of capital the financial institution is prepared to risk losing, can be expressed quantitatively. Others, such as risks to reputation, are likely to be expressed qualitatively. The risk appetite must then be converted into directives for the risks taken by business units. Appropriate limits are then set on trading risk, how concentrated the financial institution's portfolio can become, how much credit risk can be taken, and so on. The risks being taken should be continually monitored to ensure that they are consistent with the financial institution's risk appetite.

Developing a good risk culture is an important part of ERM. A key element of the risk culture is ensuring that medium- to long-term risks are considered when opportunities that offer short-term profits are evaluated. There are many examples of financial institutions aggressively pursuing short-term profits when the activity being undertaken has led several years later to serious problems such as lawsuits, fines, or a loss of reputation. The bonus structure within financial institutions can encourage employees to use the maximization of short-term profits as their sole criterion for decision making. Since the 2008 credit crisis, there have been attempts to change this by introducing deferred bonuses and clawbacks.

It is important for a financial institution to develop procedures for identifying worst-case outcomes. Unfortunately, there are many biases that are liable to cause managers to misjudge future outcomes and their probabilities of occurring. Once the risks being taken have been identified, it is important to actively manage the risks. Sometimes it will be necessary to discontinue certain activities; sometimes risks can be shared (e.g., with insurance or joint ventures); sometimes steps can be taken to reduce the impact of the risks; and sometimes risks can be accommodated within the organization's risk appetite.

Further Reading

- Fraser, J., and B. J. Simkins. *Enterprise Risk Management: Today's Leading Research and Best Practices for Tomorrow's Executives*. Hoboken, NJ: John Wiley & Sons, 2010.
- Los Angeles Times*. "Wells Fargo's Pressure-Cooker Sales Culture Comes at a Cost," December 21, 2013.
- Osborn, T. "Worst Trade of All Time Pits Santander Against Portuguese Client." *Risk* (May 2014).
- Porter, M. E. *Competitive Advantage*. New York: Free Press, 1985.

Practice Questions and Problems (Answers at End of Book)

- 24.1 Explain the difference between top-down and bottom-up approaches to risk management. Why are both necessary in ERM?
- 24.2 What is the Bowman paradox?
- 24.3 "The risk culture at a bank should lead it to consider the longer-term implications of decisions as well as short-term profits." Discuss.
- 24.4 What steps can a bank take to encourage employees to consider more than short-term profits?
- 24.5 What distinguishes enterprise risk management from more traditional approaches to risk management?
- 24.6 A fund's risk appetite is such that it wants to be 90% certain that it will not lose more than 20% in any one year. Using the performance of the S&P 500 between 1970 and 2021 (see Table 24.1), determine the beta the fund should have. Assume a risk-free rate of 3% per annum.
- 24.7 Calculate the interest rate paid by MdP six years after the beginning of the deal in Business Snapshot 24.2 if the Euribor rate proves to be 8% from year 2 onward.
- 24.8 Give three examples of cognitive biases.

Further Questions

- 24.9 Assume that in Business Snapshot 24.2 the change in the three-month Euribor rate in each quarter is normally distributed with mean zero and a standard deviation equal to x basis points. Use Monte Carlo simulation (500 trials) to calculate a probability distribution for the average interest rate paid by MdP over the four-year period for values of x equal to 10, 20, and 50.
- 24.10 A fund's risk appetite is such that it wants to be 97.5% certain it will not lose more than 25% in any one year. Using the performance of the S&P 500 between 1970 and 2021 (see Table 24.1), determine the beta the fund should have. Assume a risk-free rate of 2.5% per annum.

Part Six

REGULATION

Chapter 25

Basel I, Basel II, and Solvency II

An agreement in 1988, known as the Basel Accord, marked the start of international standards for bank regulation. Since 1988, bank regulation has been an evolutionary process. New regulations have modified previous regulations, but often approaches used in previous regulations have been preserved. In order to understand the current regulatory environment, it is therefore necessary to understand historical developments. This chapter explains the evolution of the regulatory environment prior to the 2007 Global Financial Crisis. Chapters 26 and 27 will cover developments since the crisis.

This chapter starts by explaining the 1988 Basel Accord (now known as Basel I), netting provisions, and the 1996 Amendment. It then moves on to discuss Basel II, which is a major overhaul of the regulations and was implemented by many banks throughout the world in about 2007. Finally, it reviews Solvency II, a regulatory framework for insurance companies, which is broadly similar to Basel II and was implemented by the European Union in 2016.

25.1 The Reasons for Regulating Banks

The main purpose of bank regulation is to ensure that a bank keeps enough capital for the risks it takes. It is not possible to eliminate altogether the possibility of a bank failing, but governments want to make the probability of default for any given bank very small. By

BUSINESS SNAPSHOT 25.1

Systemic Risk

Systemic risk is the risk that a default by one financial institution will create a “ripple effect” that leads to defaults by other financial institutions and threatens the stability of the financial system. There are huge numbers of over-the-counter transactions between banks. If Bank A fails, Bank B may take a huge loss on the transactions it has with Bank A. This in turn could lead to Bank B failing. Bank C that has many outstanding transactions with both Bank A and Bank B might then take a large loss and experience severe financial difficulties; and so on.

The financial system has survived defaults such as Drexel in 1990, Barings in 1995, and Lehman Brothers in 2008 very well, but regulators continue to be concerned. During the market turmoil of 2007 and 2008, many large financial institutions were bailed out, rather than being allowed to fail, because governments were concerned about systemic risk.

doing this, they hope to create a stable economic environment where private individuals and businesses have confidence in the banking system.

It is tempting to argue: “Bank regulation is unnecessary. Even if there were no regulations, banks would manage their risks prudently and would strive to keep a level of capital that is commensurate with the risks they are taking.” Unfortunately, history does not support this view. There is little doubt that regulation has played an important role in increasing bank capital and making banks more aware of the risks they are taking.

As discussed in Section 2.3, governments provide deposit insurance programs to protect depositors. Without deposit insurance, banks that took excessive risks relative to their capital base would find it difficult to attract deposits. However, the impact of deposit insurance is to create an environment where depositors are less discriminating. A bank can take large risks without losing its deposit base.¹ The last thing a government wants is to create a deposit insurance program that results in banks taking more risks. It is therefore essential that deposit insurance be accompanied by regulation concerned with capital requirements.

A major concern of governments is what is known as *systemic risk*. This is the risk that a failure by a large bank will lead to failures by other large banks and a collapse of the financial system. The way this can happen is described in Business Snapshot 25.1. When a bank or other large financial institution does get into financial difficulties, governments

¹As explained in Chapter 3, this is an example of what insurance companies term moral hazard. The existence of an insurance contract changes the behavior of the insured party.

have a difficult decision to make. If they allow the financial institution to fail, they are putting the financial system at risk. If they bail out the financial institution, they are sending the wrong signals to the market. There is a danger that large financial institutions will be less vigilant in controlling risks if they know that they are “too big to fail” and the government will always bail them out.

During the market turmoil of 2008, the decision was taken to bail out some financial institutions in the United States and Europe. However, Lehman Brothers was allowed to fail in September 2008. Possibly, the United States government wanted to make it clear to the market that bailouts for large financial institutions were not automatic. However, the decision to let Lehman Brothers fail has been criticized because arguably it made the Global Financial Crisis worse.

25.2 Bank Regulation Pre-1988

Prior to 1988, bank regulators within a country tended to regulate bank capital by setting minimum levels for the ratio of capital to total assets. However, definitions of capital and the ratios considered acceptable varied from country to country. Some countries enforced their regulations more diligently than other countries. Increasingly, banks were competing globally and a bank operating in a country where capital regulations were lax was considered to have a competitive edge over one operating in a country with tighter, more strictly enforced, capital regulations. In addition, the huge exposures created by loans from the major international banks to less developed countries such as Mexico, Brazil, and Argentina, as well as the accounting games sometimes used for those exposures (see Business Snapshot 2.3) were starting to raise questions about the adequacy of capital levels.

Another problem was that the types of transactions entered into by banks were becoming more complicated. The over-the-counter derivatives market for products such as interest rate swaps, currency swaps, and foreign exchange options was growing fast. These contracts increase the credit risks being taken by a bank. Consider, for example, an interest rate swap. If the counterparty in the interest rate swap transaction defaults when the swap has a positive value to the bank and a negative value to the counterparty, the bank is liable to lose money. The potential future exposure from derivatives was not reflected in the bank's reported assets. As a result, it had no effect on the level of assets reported by a bank and therefore no effect on the amount of capital the bank was required to keep. It became apparent to regulators that the value of total assets was no longer a good indicator of the total risks being taken. A more sophisticated approach than that of setting minimum levels for the ratio of capital to total balance-sheet assets was needed.

The Basel Committee was formed in 1974. The committee consisted of representatives from Belgium, Canada, France, Germany, Italy, Japan, Luxembourg, the Netherlands, Sweden, Switzerland, the United Kingdom, and the United States. It met regularly in Basel, Switzerland, under the patronage of the Bank for International

Settlements. The first major result of these meetings was a document titled “International Convergence of Capital Measurement and Capital Standards.” This was referred to as “The 1988 BIS Accord” or just “The Accord.” Later it became known as Basel I.

25.3 The 1988 BIS Accord

The 1988 BIS Accord was the first attempt to set international risk-based standards for capital adequacy. It has been subject to much criticism as being too simple and somewhat arbitrary. In fact, the Accord was a huge achievement. It was signed by all 12 members of the Basel Committee and paved the way for significant increases in the resources banks devote to measuring, understanding, and managing risks. The key innovation in the 1988 Accord was the Cooke ratio.

25.3.1 The Cooke Ratio

The Cooke ratio² considers credit risk exposures that are both on-balance-sheet and off-balance-sheet. It is based on what is known as the bank’s total *risk-weighted assets* (also sometimes referred to as the *risk-weighted amount*). This is a measure of the bank’s total credit exposure.

Credit risk exposures can be divided into three categories:

1. Those arising from on-balance sheet assets (excluding derivatives)
2. Those arising for off-balance sheet items (excluding derivatives)
3. Those arising from over-the-counter derivatives

Consider the first category. Each on-balance-sheet asset is assigned a risk weight reflecting its credit risk. A sample of the risk weights specified in the Accord is shown in Table 25.1. Cash and securities issued by governments of OECD countries (members of the Organisation for Economic Co-operation and Development) are considered to have virtually zero risk and have a risk weight of zero. Loans to corporations have a risk weight of 100%. Loans to banks and government agencies in OECD countries have a risk weight of 20%. Uninsured residential mortgages have a risk weight of 50%. The total of the risk-weighted assets for N on-balance-sheet items equals

$$\sum_{i=1}^N w_i L_i$$

where L_i is the principal amount of the i th item and w_i is its risk weight.

²The ratio is named after Peter Cooke, from the Bank of England.

Table 25.1 Risk Weights for On-Balance-Sheet Items

Risk Weight (%)	Asset Category
0	Cash, gold bullion, claims on OECD governments such as Treasury bonds or insured residential mortgages
20	Claims on OECD banks and OECD public-sector entities such as securities issued by U.S. government agencies or claims on municipalities
50	Uninsured residential mortgage loans
100	All other claims such as corporate bonds and less developed country debt, claims on non-OECD banks

Example 25.1

The assets of a bank consist of \$100 million of corporate loans, \$10 million of OECD government bonds, and \$50 million of residential mortgages. The total of the risk-weighted assets is

$$1.0 \times 100 + 0.0 \times 10 + 0.5 \times 50 = 125$$

or \$125 million.

Consider next the second category. This includes bankers' acceptances, guarantees, and loan commitments. A credit equivalent amount is calculated by applying a conversion factor to the principal amount of the instrument. Instruments that from a credit perspective are considered to be similar to loans, such as bankers' acceptances, have a conversion factor of 100%. Others, such as note issuance facilities (where a bank agrees that a company can issue short-term paper on pre-agreed terms in the future), have lower conversion factors.

Finally, consider the third category. For an over-the-counter derivative such as an interest rate swap or a forward contract, the credit equivalent amount is calculated as

$$\max(V, 0) + aL \quad (25.1)$$

where V is the current value of the derivative to the bank, a is an *add-on factor*, and L is the principal amount. The first term in equation (25.1) is the current exposure. If the counterparty defaults today and V is positive, the contract is an asset to the bank and the bank is liable to lose V . If the counterparty defaults today and V is negative, the contract is an asset to the counterparty and there will be neither a gain nor a loss to the bank. The bank's exposure is therefore $\max(V, 0)$. (More details on the bilaterally cleared over-the-counter derivatives market are in Chapters 6 and 18. At this stage, netting was not recognized in the calculation of bank required capital.) The add-on amount, aL , is an allowance for the possibility of the exposure increasing in the future. Examples of the add-on factor, a , are shown in Table 25.2. Equation (25.1) is known as the *current exposure method* (CEM). The add-on factors were revised and extended in the years following 1988.

Table 25.2 Add-On Factors as a Percent of Principal for Derivatives

Remaining Maturity (yr)	Interest Rate	Exchange Rate and Gold	Equity	Precious Metals Except Gold	Other Commodities
< 1	0.0	1.0	6.0	7.0	10.0
1 to 5	0.5	5.0	8.0	7.0	12.0
> 5	1.5	7.5	10.0	8.0	15.0

Example 25.2

A bank has entered into a \$100 million interest rate swap with a remaining life of four years. The current value of the swap is \$2.0 million. In this case, the add-on amount is 0.5% of the principal so that the credit equivalent amount is \$2.0 million plus \$0.5 million or \$2.5 million.

The credit equivalent amount arising from either the second or third category of exposures is multiplied by the risk weight for the counterparty in order to calculate the risk-weighted assets. The risk weights are similar to those in Table 25.1 except that the risk weight for a corporation is 0.5 rather than 1.0.

Example 25.3

Consider again the bank in Example 25.2. If the interest rate swap is with a corporation, the risk-weighted assets are 2.5×0.5 or \$1.25 million. If it is with an OECD bank, the risk-weighted assets are 2.5×0.2 or \$0.5 million.

Putting all this together, the total risk-weighted assets for a bank with N on-balance-sheet items and M off-balance-sheet items is

$$\sum_{i=1}^N w_i L_i + \sum_{j=1}^M w_j^* C_j \quad (25.2)$$

Here, L_i is the principal of the i th on-balance-sheet asset and w_i is the risk weight for the asset; C_j is the credit equivalent amount for the j th derivative or other off-balance-sheet item and w_j^* is the risk weight of the counterparty for this j th item.

25.3.2 Capital Requirement

The Accord required banks to keep capital equal to at least 8% of the risk-weighted assets. The capital had two components:

1. *Tier 1 Capital*. This consists of items such as equity and noncumulative perpetual preferred stock.³ (Goodwill is subtracted from equity.⁴)
2. *Tier 2 Capital*. This is sometimes referred to as Supplementary Capital. It includes instruments such as cumulative perpetual preferred stock,⁵ certain types of 99-year debenture issues, and subordinated debt (i.e., debt subordinated to depositors) with an original life of more than five years.

Equity capital is the most important type of capital because it absorbs losses. If equity capital is greater than losses, a bank can continue as a going concern. If equity capital is less than losses, the bank is insolvent. In the latter case, Tier 2 capital becomes relevant. Because it is subordinate to depositors, it provides a cushion for depositors. If a bank is wound up after its Tier I capital has been used up, losses should be borne first by the Tier 2 capital and, only if that is insufficient, by depositors. (See Section 2.2.)

The Accord required at least 50% of the required capital (that is, 4% of the risk-weighted assets) to be in Tier 1. Furthermore, the Accord required 2% of risk-weighted assets to be common equity. (Note that the Basel Committee has updated its definition of instruments that are eligible for Tier 1 capital and its definition of common equity in Basel III.)

The bank supervisors in some countries require banks to hold more capital than the minimum specified by the Basel Committee and some banks themselves have a target for the capital they will hold that is higher than that specified by their bank supervisors.

25.4 The G-30 Policy Recommendations

In 1993, a working group consisting of end-users, dealers, academics, accountants, and lawyers involved in derivatives published a report that contained 20 risk management recommendations for dealers and end-users of derivatives and four recommendations for legislators, regulators, and supervisors. The report was based on a detailed survey of 80 dealers and 72 end-users worldwide. The survey involved both questionnaires and in-depth interviews. The report is not a regulatory document, but it has been influential in the development of risk management practices.

A brief summary of the important recommendations is as follows:

1. A company's policies on risk management should be clearly defined and approved by senior management, ideally at the board of directors level. Managers at all levels should enforce the policies.

³Noncumulative perpetual preferred stock is preferred stock lasting forever where there is a predetermined dividend rate. Unpaid dividends do not cumulate (that is, the dividends for one year are not carried forward to the next year).

⁴Goodwill is created when one company acquires another. It is the purchase price minus the book value of the assets. It represents the intangible assets acquired.

⁵In cumulative preferred stock, unpaid dividends cumulate. Any backlog of dividends must be paid before dividends are paid on the common stock.

2. Derivatives positions should be marked to market (i.e., revalued using a model that is consistent with market prices) at least once a day.
3. Derivatives dealers should measure market risk using a consistent measure such as value at risk. Limits to the market risks that are taken should be set.
4. Derivatives dealers should carry out stress tests to determine potential losses under extreme market conditions.
5. The risk management function should be set up so that it is independent of the trading operation.
6. Credit exposures arising from derivatives trading should be assessed based on the current replacement value of existing positions and potential future replacement costs.
7. Credit exposures to a counterparty should be aggregated in a way that reflects enforceable netting agreements. (We talk about netting in the next section.)
8. The individuals responsible for setting credit limits should be independent of those involved in trading.
9. Dealers and end-users should assess carefully both the costs and benefits of credit risk mitigation techniques such as collateralization and downgrade triggers. In particular, they should assess their own capacity and that of their counterparties to meet the cash flow requirement of downgrade triggers. (Downgrade triggers are discussed in Chapter 18.)
10. Only individuals with the appropriate skills and experience should be allowed to have responsibility for trading derivatives, supervising the trading, carrying out back office functions in relation to the trading, and so on.
11. There should be adequate systems in place for data capture, processing, settlement, and management reporting.
12. Dealers and end-users should account for the derivatives transactions used to manage risks so as to achieve a consistency of income recognition treatment between those instruments and the risks being managed.

25.5 Netting

Participants in the over-the-counter derivatives market have traditionally signed an International Swaps and Derivatives Association (ISDA) master agreement covering their derivatives trades. The word *netting* refers to a clause in the master agreement, which states that in the event of a default all transactions are considered as a single transaction. Effectively, this means that, if a company defaults on one transaction that is covered by the master agreement, it must default on all transactions covered by the master agreement.

Netting and ISDA master agreements were discussed in Chapters 6 and 18. Netting can have the effect of substantially reducing credit risk. Consider a bank that has three swap transactions outstanding with a particular counterparty. The transactions are worth +\$24 million, −\$17 million, and +\$8 million to the bank. Suppose that the counterparty experiences financial difficulties and defaults on its outstanding obligations.

To the counterparty the three transactions have values of $-\$24$ million, $+\$17$ million, and $-\$8$ million, respectively. Without netting, the counterparty would default on the first transaction, keep the second transaction, and default on the third transaction. Assuming no recovery, the loss to the bank would be $\$32$ ($= 24 + 8$) million. With netting, the counterparty is required to default on the second transaction as well. The loss to the bank is then $\$15$ ($= 24 - 17 + 8$) million.

More generally, suppose that a financial institution has a portfolio of N derivatives outstanding with a particular counterparty and that the current value of the i th derivative is V_i . Without netting, the financial institution's exposure in the event of a default today is

$$\sum_{i=1}^N \max(V_i, 0)$$

With netting, it is

$$\max\left(\sum_{i=1}^N V_i, 0\right)$$

Without netting, the exposure is the payoff from a portfolio of options. With netting, the exposure is the payoff from an option on a portfolio.

The 1988 Basel Accord did not take netting into account in setting capital requirements. From equation (25.1) the credit equivalent amount for a portfolio of derivatives with a counterparty under the Accord was

$$\sum_{i=1}^N [\max(V_i, 0) + a_i L_i]$$

where a_i is the add-on factor for the i th transaction and L_i is the principal for the i th transaction.

By 1995, netting had been successfully tested in the courts in many jurisdictions. As a result, the 1988 Accord was modified to allow banks to reduce their credit equivalent totals when enforceable bilateral netting agreements were in place. The first step was to calculate the net replacement ratio, NRR. This is the ratio of the current exposure with netting to the current exposure without netting:

$$\text{NRR} = \frac{\max\left(\sum_{i=1}^N V_i, 0\right)}{\sum_{i=1}^N \max(V_i, 0)}$$

The credit equivalent amount was modified to

$$\max\left(\sum_{i=1}^N V_i, 0\right) + (0.4 + 0.6 \times \text{NRR}) \sum_{i=1}^N a_i L_i$$

Table 25.3 Portfolio of Derivatives with a Particular Counterparty

Transaction	Principal, L_i	Current Value, V_i	Add-On Amount, $a_i L_i$
3-year interest rate swap	1,000	-60	5
6-year foreign exchange forward	1,000	70	75
9-month option on a stock	500	55	30

Example 25.4

Consider the example in Table 25.3, which shows a portfolio of three derivatives that a bank has with a particular counterparty. The third column shows the current mark-to-market values of the transactions and the fourth column shows the add-on amount calculated from Table 25.2. The current exposure with netting is $-60 + 70 + 55 = 65$. The current exposure without netting is $0 + 70 + 55 = 125$. The net replacement ratio is given by

$$\text{NRR} = \frac{65}{125} = 0.52$$

The total of the add-on amounts, $\sum a_i L_i$, is $5 + 75 + 30 = 110$. The credit equivalent amount when netting agreements are in place is $65 + (0.4 + 0.6 \times 0.52) \times 110 = 143.32$. Without netting, the credit equivalent amount is $125 + 110 = 235$. Suppose that the counterparty is an OECD bank so that the risk weight is 0.2. This means that the risk-weighted assets with netting is $0.2 \times 143.32 = 28.66$. Without netting, it is $0.2 \times 235 = 47$.

25.6 The 1996 Amendment

In 1995, the Basel Committee issued a consultative proposal to amend the 1988 Accord. This became known as the “1996 Amendment.” It was implemented in 1998 and was then sometimes referred to as “BIS 98.” The amendment involves keeping capital for the market risks associated with trading activities.

Marking to market is the practice of revaluing assets and liabilities daily using a model that is calibrated to current market prices. It is also known as *fair value accounting*. Banks are required to use fair value accounting for all assets and liabilities that are held for trading purposes. This includes most derivatives, marketable equity securities, foreign currencies, and commodities. These items constitute what is referred to as the bank’s *trading book*.

Under the 1996 Amendment, the credit risk capital charge in the 1988 Accord continued to apply to all on-balance-sheet and off-balance-sheet items in the trading and banking book, except positions in the trading book that consisted of (a) debt and equity traded securities and (b) positions in commodities and foreign exchange. The

Amendment introduced a capital charge for the market risk associated with all items in the trading book.⁶

The 1996 Amendment outlined a standardized approach for measuring the capital charge for market risk. The standardized approach assigned capital separately to each of debt securities, equity securities, foreign exchange risk, commodities risk, and options. No account was taken of correlations between different types of instruments. The more sophisticated banks with well-established risk management functions were allowed to use an “internal model-based approach” for setting market risk capital. This involved calculating a value-at-risk measure and converting it into a capital requirement using a formula specified in the 1996 Amendment. Most large banks preferred to use the internal model-based approach because it better reflected the benefits of diversification and led to lower capital requirements.

The value-at-risk measure used in the internal model-based approach was calculated with a 10-day time horizon and a 99% confidence level. It is the loss that has a 1% chance of being exceeded over a 10-day period. The capital requirement is

$$\max(\text{VaR}_{t-1}, m_c \times \text{VaR}_{\text{avg}}) + \text{SRC} \quad (25.3)$$

where m_c is a multiplicative factor, and SRC is a *specific risk charge*. The variable VaR_{t-1} is the previous day’s value at risk and VaR_{avg} is the average value at risk over the past 60 days. The minimum value for m_c is 3. Higher values may be chosen by regulators for a particular bank if tests reveal inadequacies in the bank’s value-at-risk model, as will be explained shortly.

The first term in equation (25.3) covers risks relating to movements in broad market variables such as interest rates, exchange rates, stock indices, and commodity prices. The second term, SRC, covers risks related to specific companies such as those concerned with movements in a company’s stock price or changes in a company’s credit spread.

Consider the first term and assume $m_c = 3$. In most circumstances, the most recently calculated VaR, VaR_{t-1} , is less than three times the average VaR over the past 60 days. This gives the capital requirement for movements in broad market variables as simply

$$3 \times \text{VaR}_{\text{avg}}$$

The most popular method for calculating VaR is historical simulation, described in Chapter 12. As explained in Chapter 11, banks almost invariably calculate a one-day 99% VaR in the first instance. When formulating the 1996 amendment, regulators explicitly stated that the 10-day 99% VaR can be calculated as $\sqrt{10}$ times the one-day 99% VaR. This means that, when the capital requirement for a bank is calculated as m_c times the average 10-day 99% VaR, it is to all intents and purposes $m_c \times \sqrt{10} = 3.16m_c$ times the average one-day 99% VaR. If $m_c = 3$, it is 9.48 times the average one-day VaR.

Consider next SRC. One security that gives rise to an SRC is a corporate bond. There are two components to the risk of this security: interest rate risk and credit risk of

⁶Certain nontrading book positions that are used to hedge positions in the trading book can be included in the calculation of the market risk capital charge.

the corporation issuing the bond. The interest rate risk is captured by the first term in equation (25.3); the credit risk is captured by the SRC.⁷ The 1996 Amendment proposed standardized methods for assessing the SRC, but allowed banks to use internal models once regulatory approval for the models had been obtained.

The internal model for SRC must involve calculating a 10-day 99% value at risk for specific risks. Regulators calculate capital by applying a multiplicative factor (similar to m_c) to the value at risk. This multiplicative factor must be at least 4 and the resultant capital must be at least 50% of the capital given by the standardized approach. An approach for calculating SRC for credit-spread risk was indicated in Section 19.5.

The total capital a bank was required to keep after the implementation of the 1996 Amendment was the sum of (a) credit risk capital equal to 8% of the risk-weighted assets (RWA) and (b) market risk capital as explained in this section. For convenience, an RWA for market risk capital was defined as 12.5 multiplied by the amount given in equation (25.3). This means that the total capital required for credit and market risk is given by

$$\text{Total Capital} = 0.08 \times (\text{credit risk RWA} + \text{market risk RWA}) \quad (25.4)$$

A bank had more flexibility in the type of capital it used for market risk. It could use Tier 1 or Tier 2 capital. It could also use what is termed Tier 3 capital. This consists of short-term subordinated debt with an original maturity of at least two years that is unsecured and fully paid up. (Tier 3 capital was eliminated under Basel III.)

25.6.1 Back-Testing

The BIS Amendment required the one-day 99% VaR that a bank calculates to be back-tested over the previous 250 days. As described in Section 11.10, this involves using the bank's current procedure for estimating VaR for each of the most recent 250 days. If the actual loss that occurred on a day is greater than the VaR level calculated for the day, an "exception" is recorded. Calculations are carried out (a) including changes that were made to the portfolio on the day being considered and (b) assuming that no changes were made to the portfolio on the day being considered. (Regulators like to see both calculations.)

If the number of exceptions during the previous 250 days is less than 5, m_c is normally set equal to 3. If the number of exceptions is 5, 6, 7, 8, and 9, the value of the m_c is set equal to 3.4, 3.5, 3.65, 3.75, and 3.85, respectively. The bank supervisor has some discretion as to whether the higher multipliers are used. They will normally be applied when the reason for the exceptions is identified as a deficiency in the VaR model being used. If changes in the bank's positions during the day result in exceptions, the higher multiplier should be considered, but does not have to be used. When the only reason that is identified is bad luck, no guidance is provided for the supervisor. In circumstances

⁷As mentioned earlier, the 1988 credit risk capital charge did not apply to debt securities in the trading book under the 1996 Amendment.

where the number of exceptions is 10 or more, the Basel Amendment requires the multiplier to be set at 4. Problem 25.18 considers these guidelines in the context of the statistical tests we discussed in Chapter 11.

25.7 Basel II

The 1988 Basel Accord improved the way capital requirements were determined, but it does have significant weaknesses. Under the Accord, all loans by a bank to a corporation have a risk weight of 100% and require the same amount of capital. A loan to a corporation with a AAA credit rating is treated in the same way as one to a corporation with a B credit rating.⁸ Also, in Basel I there was no model of default correlation.

In June 1999, the Basel Committee proposed new rules that have become known as Basel II. These were revised in January 2001 and April 2003. A number of Quantitative Impact Studies (QISs) were carried out prior to the implementation of the new rules to test them by calculating the amount of capital that would be required if the rules had been in place.⁹ A final set of rules agreed to by all members of the Basel Committee was published in June 2004. This was updated in November 2005. Implementation of the rules began in 2007 after a further QIS.

The Basel II capital requirements applied to “internationally active” banks. In the United States, there are many small regional banks and the U.S. regulatory authorities decided that Basel II would not apply to them. (These banks are regulated under what is termed Basel IA, which is similar to Basel I.) In Europe, all banks, large or small, were regulated under Basel II. Furthermore, the European Union required the Basel II rules to be applied to securities companies as well as banks.

The Basel II is based on three “pillars”:

1. Minimum Capital Requirements
2. Supervisory Review
3. Market Discipline

In Pillar 1, the minimum capital requirement for credit risk in the banking book is calculated in a new way that reflects the credit risk of counterparties. The capital requirement for market risk remains unchanged from the 1996 Amendment and there is a new capital charge for operational risk. The general requirement in Basel I that banks hold a total capital equal to 8% of risk-weighted assets (RWA) remains unchanged. When the capital requirement for a particular risk is calculated directly rather than in a way involving RWAs, it is multiplied by 12.5 to convert it into an RWA-equivalent. As a result it is always the case that

⁸Credit ratings are discussed in Section 1.7.

⁹One point to note about the QISs is that they do not take account of changes banks may choose to make to their portfolios to minimize their capital requirements once the new rules have been implemented.

$$\begin{aligned}\text{Total Capital} &= 0.08 \times (\text{credit risk RWA} + \text{market risk RWA} \\ &\quad + \text{operational risk RWA})\end{aligned}\tag{25.5}$$

Pillar 2 is concerned with the supervisory review process. It covers both quantitative and qualitative aspects of the ways risk is managed within a bank. Supervisors are required to ensure that a bank has a process in place for ensuring that capital levels are maintained. Banks are expected to keep more than the minimum regulatory capital to allow for fluctuations in capital requirements and possible difficulties in raising capital at short notice. Regulators in different countries are allowed some discretion in how rules are applied (so that they can take account of local conditions), but overall consistency in the application of the rules is required. Pillar 2 places more emphasis on early intervention when problems arise. Supervisors are required to do far more than just ensuring that the minimum capital required under Basel II is held. Part of their role is to encourage banks to develop and use better risk management techniques and to evaluate those techniques. They should evaluate risks that are not covered by Pillar 1 (e.g., concentration risks) and enter into an active dialogue with banks when deficiencies are identified.

The third pillar, market discipline, requires banks to disclose more information about the way they allocate capital and the risks they take. The idea here is that banks will be subjected to added pressure to make sound risk management decisions if shareholders and potential shareholders have more information about those decisions.

25.8 Credit Risk Capital Under Basel II

For credit risk, Basel II specified three approaches:

1. The Standardized Approach
2. The Foundation Internal Ratings Based (IRB) Approach
3. The Advanced IRB Approach

However, the United States (which, as mentioned earlier, chose to apply Basel II only to large banks) decided that only the IRB approach can be used.

25.8.1 *The Standardized Approach*

The standardized approach is used by banks that are not sufficiently sophisticated (in the eyes of the regulators) to use the internal ratings approaches. The standardized approach is similar to Basel I except for the calculation of risk weights.¹⁰ Some of the

¹⁰Ratios calculated using the new weights are sometimes referred to as McDonough ratios after William McDonough, the head of the Basel Committee.

Table 25.4 Risk Weights as a Percent of Principal for Exposures to Countries, Banks, and Corporations Under Basel II's Standardized Approach

	AAA to AA–	A+ to A–	BBB+ to BBB–	BB+ to BB–	B+ to B–	Below B–	Unrated
Country*	0	20	50	100	100	150	100
Banks**	20	50	50	100	100	150	50
Corporations	20	50	100	100	150	150	100

*Includes exposures to the country's central bank.

**National supervisors have options as outlined in the text.

new rules here are summarized in Table 25.4. Comparing Table 25.4 with Table 25.1, we see that the OECD status of a bank or a country is no longer considered important under Basel II. The risk weight for a country (sovereign) exposure ranges from 0% to 150% and the risk weight for an exposure to another bank or a corporation ranges from 20% to 150%. In Table 25.1, OECD banks were implicitly assumed to be lesser credit risks than corporations. An OECD bank attracted a risk weight of 20% while a corporation attracted a risk weight of 100%. Table 25.4 treats banks and corporations much more equitably. An interesting observation from Table 25.4 for a country, corporation, or bank that wants to borrow money is that it may be better to have no credit rating at all than a very poor credit rating. (Usually a company gets a credit rating when it issues a publicly traded bond.) Supervisors are allowed to apply lower risk weights (20% rather than 50%, 50% rather than 100%, and 100% rather than 150%) when exposures are to the country in which the bank is incorporated or to that country's central bank.

The risk weight for mortgages in the Basel II standardized approach was 35% and that for other retail loans was 75%. For claims on banks, the rules are somewhat complicated. Instead of using the risk weights in Table 25.4, national supervisors can choose to base capital requirements on the rating of the country in which the bank is incorporated. The risk weight assigned to the bank will be 20% if the country of incorporation has a rating between AAA and AA–, 50% if it is between A+ and A–, 100% if it is between BBB+ and B–, 150% if it is below B–, and 100% if it is unrated. Another complication is that, if national supervisors elect to use the rules in Table 25.4, they can choose to treat claims with a maturity less than three months more favorably so that the risk weights are 20% if the rating is between AAA+ and BBB–, 50% if it is between BB+ and B–, 150% if it is below B–, and 20% if it is unrated.

In December 2017, the Basel Committee created a revised, more granular standardized approach for determining risk weights for credit exposures. It also limited the use of the Advanced IRB approaches in some situations. As explained in Chapter 26, the revised standardized approach provides a floor for the determination of total capital requirements.

Example 25.5

Suppose that the assets of a bank consist of \$100 million of loans to corporations rated A, \$10 million of government bonds rated AAA, and \$50 million of residential mortgages. Under the Basel II standardized approach, the total of the risk-weighted assets is

$$0.5 \times 100 + 0.0 \times 10 + 0.35 \times 50 = 67.5$$

or \$67.5 million. This compares with \$125 million under Basel I. (See Example 25.1.)

25.8.2 Adjustments for Collateral

There are two ways banks can adjust risk weights for collateral under Basel II. The first is termed the *simple approach* and is similar to an approach used in Basel I. The second is termed the *comprehensive approach*. Banks have a choice as to which approach is used in the banking book, but must use the comprehensive approach to calculate capital for counterparty credit risk in the trading book.

Under the simple approach, the risk weight of the counterparty is replaced by the risk weight of the collateral for the part of the exposure covered by the collateral. (The exposure is calculated after netting.) For any exposure not covered by the collateral, the risk weight of the counterparty is used. The minimum level for the risk weight applied to the collateral is 20%.¹¹ A requirement is that the collateral must be revalued at least every six months and must be pledged for at least the life of the exposure.

Under the comprehensive approach, banks adjust the size of their exposure upward to allow for possible increases in the exposure and adjust the value of the collateral downward to allow for possible decreases in the value of the collateral.¹² (The adjustments depend on the volatility of the exposure and the collateral.) A new exposure equal to the excess of the adjusted exposure over the adjusted value of the collateral is calculated and the counterparty's risk weight is applied to this exposure. The adjustments applied to the exposure and the collateral can be calculated using rules specified in Basel II or, with regulatory approval, using a bank's internal models. Where netting arrangements apply, exposures and collateral are separately netted and the adjustments made are weighted averages.

Example 25.6

Suppose that an \$80 million exposure to a particular counterparty is secured by collateral worth \$70 million. The collateral consists of bonds issued by an A-rated company. The

¹¹An exception is when the collateral consists of cash or government securities with the currency of the collateral being the same as the currency of the exposure.

¹²An adjustment to the exposure is not likely to be necessary on a loan, but is likely to be necessary on an over-the-counter derivative. The adjustment is in addition to the add-on factor.

counterparty has a rating of B+. The risk weight for the counterparty is 150% and the risk weight for the collateral is 50%. The risk-weighted assets applicable to the exposure using the simple approach is

$$0.5 \times 70 + 1.50 \times 10 = 50$$

or \$50 million.

Consider next the comprehensive approach. Assume that the adjustment to exposure to allow for possible future increases in the exposure is +10% and the adjustment to the collateral to allow for possible future decreases in its value is -15%. The new exposure is

$$1.1 \times 80 - 0.85 \times 70 = 28.5$$

or \$28.5 million, and a risk weight of 150% is applied to this exposure to give risk-adjusted assets equal to \$42.75 million.

25.8.3 The IRB Approach

The model underlying the IRB approach is shown in Figure 25.1. Regulators base the capital requirement on the value at risk calculated using a one-year time horizon and a 99.9% confidence level. They recognize that expected losses are usually covered by the way a financial institution prices its products. (For example, the interest charged by a bank on a loan portfolio is designed to recover expected loan losses on the portfolio.) The capital required is therefore the value at risk minus the expected loss.

The value at risk is calculated using the one-factor Gaussian copula model of time to default that we discussed in Section 9.6. Assume that a bank has a very large number of obligors and the i th obligor has a one-year probability of default equal to PD_i . The copula correlation between each pair of obligors is ρ .¹³ As in Section 9.6, we define

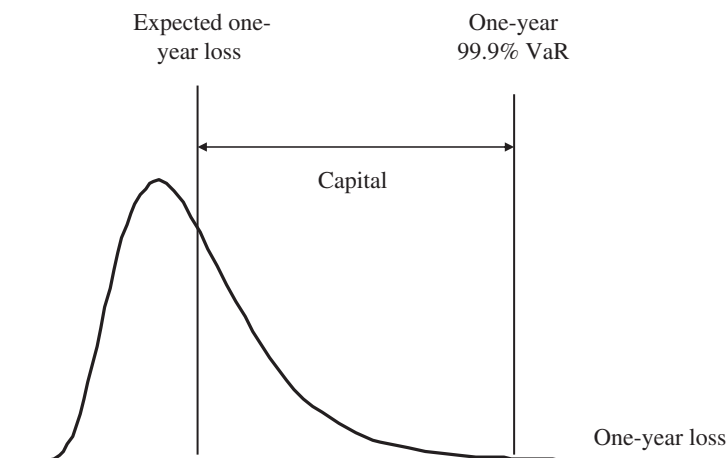


Figure 25.1 The Loss Probability Density Function and the Capital Required by a Financial Institution

¹³Note that the Basel Committee publications use R , not ρ , to denote the copula correlation.

$$\text{WCDR}_i = N \left[\frac{N^{-1}(\text{PD}_i) + \sqrt{\rho} N^{-1}(0.999)}{\sqrt{1 - \rho}} \right] \quad (25.6)$$

where WCDR_i denotes the “worst-case default rate” defined so that the bank is 99.9% certain it will not be exceeded next year for the i th counterparty. Gordy’s (2003) research shows that for a large portfolio of instruments (loans, loan commitments, derivatives, and so on) that have the same ρ , in a one-factor model the one-year 99.9% VaR is approximately¹⁴

$$\sum_i \text{EAD}_i \times \text{LGD}_i \times \text{WCDR}_i$$

where EAD_i is the exposure at default of the i th counterparty and LGD_i is the loss given default for the i th counterparty.

The variable EAD_i is the dollar amount that is expected to be owed by the i th counterparty at the time of default during the next year. The variable LGD_i is the proportion of EAD_i that is expected to be lost in the event of default. For example, if a bank expects to recover 30% of the amount owed in the event of default, $\text{LGD}_i = 0.7$.

The expected loss from defaults is

$$\sum_i \text{EAD}_i \times \text{LGD}_i \times \text{PD}_i$$

The capital required in Figure 25.1 is the excess of the 99.9% worst-case loss over the expected loss. It is therefore

$$\sum_i \text{EAD}_i \times \text{LGD}_i \times (\text{WCDR}_i - \text{PD}_i) \quad (25.7)$$

We now drop the subscripts and define for a counterparty:

PD: The probability that the counterparty will default within one year (expressed as a decimal)

EAD: The exposure at default (in dollars)

LGD: The loss given default or the proportion of the exposure that is lost if there is a default (expressed as a decimal)

Table 25.5 shows how WCDR depends on PD and ρ in the Gaussian copula model. When the correlation ρ is zero, $\text{WCDR} = \text{PD}$ because in that case there is no default correlation and the percentage of loans defaulting can be expected to be the same in all years. As ρ increases, WCDR increases.

¹⁴See M. B. Gordy, “A Risk-Factor Model Foundation for Ratings-Based Bank Capital Ratios,” *Journal of Financial Intermediation* 12 (2003): 199–232.

Table 25.5 Dependence of One-Year 99.9% WCDR on PD and ρ

	PD = 0.1%	PD = 0.5%	PD = 1%	PD = 1.5%	PD = 2.0%
$\rho = 0.0$	0.1%	0.5%	1.0%	1.5%	2.0%
$\rho = 0.2$	2.8%	9.1%	14.6%	18.9%	22.6%
$\rho = 0.4$	7.1%	21.1%	31.6%	39.0%	44.9%
$\rho = 0.6$	13.5%	38.7%	54.2%	63.8%	70.5%
$\rho = 0.8$	23.3%	66.3%	83.6%	90.8%	94.4%

25.8.4 Exposure at Default

In the Foundation IRB approach, the calculation of EAD for derivatives is usually based on the CEM approach of Basel I.¹⁵ In the Advanced IRB approach, banks can use their internal models to calculate EAD.

The first stage when internal models are used is to calculate expected exposure (EE) to each counterparty at a number of future times. (This generally involves Monte Carlo simulation.) The Effective Expected Exposure (Effective EE) at a future time, t , is the maximum EE between time zero and time t . EAD is set equal to 1.4 times the Effective Expected Positive Exposure, which is the average of the Effective EEs over the first year.¹⁶

25.8.5 Corporate, Sovereign, and Bank Exposures

In the case of corporate, sovereign, and bank exposures, Basel II assumes a relationship between the correlation parameter, ρ , and the probability of default, PD, based on empirical research.¹⁷ The formula is

$$\rho = 0.12 \frac{1 - \exp(-50 \times \text{PD})}{1 - \exp(-50)} + 0.24 \left[1 - \frac{1 - \exp(-50 \times \text{PD})}{1 - \exp(-50)} \right]$$

Because $\exp(-50)$ is a very small number, this formula is to all intents and purposes

$$\rho = 0.12(1 + e^{-50 \times \text{PD}}) \quad (25.8)$$

As PD increases, ρ decreases. The reason usually given for this inverse relationship is as follows. As a company becomes less creditworthy, its PD increases and its probability of default becomes more idiosyncratic and less affected by overall market conditions.

Combining equation (25.8) with equation (25.6), we obtain the relationship between WCDR and PD in Table 25.6. WCDR is, as one would expect, an increasing function of

¹⁵As will be discussed in Section 26.4, this was revised under a regulatory change known as SA-CCR.

¹⁶See Basel Committee on Banking Supervision, "The Application of Basel II to Trading Activities and the Treatment of Double Default Effects," July 2005.

¹⁷See J. Lopez, "The Empirical Relationship between Average Asset Correlation, Firm Probability of Default, and Asset Size," *Journal of Financial Intermediation* 13, no. 2 (2004): 265–283.

Table 25.6 Relationship between One-Year 99.9% WCDR and PD for Corporate, Sovereign, and Bank Exposures

PD	0.1%	0.5%	1%	1.5%	2.0%
WCDR	3.4%	9.8%	14.0%	16.9%	19.0%

PD. However, it does not increase as fast as it would if ρ were assumed to be independent of PD.

The formula for the capital required for the counterparty is

$$\text{EAD} \times \text{LGD} \times (\text{WCDR} - \text{PD}) \times \text{MA} \tag{25.9}$$

The meaning of the first three terms in this expression should be clear from our earlier discussion leading to equation (25.7). The variable MA is the maturity adjustment and is defined as

$$\text{MA} = \frac{1 + (\text{M} - 2.5) \times b}{1 - 1.5 \times b} \tag{25.10}$$

where

$$b = [0.11852 - 0.05478 \times \ln(\text{PD})]^2$$

and M is the maturity of the exposure. The maturity adjustment is designed to allow for the fact that, if an instrument lasts longer than one year, there is a one-year credit exposure arising from a possible decline in the creditworthiness of the counterparty as well as from a possible default by the counterparty. (Note that, when $\text{M} = 1$, MA is 1.0 and has no effect.) As mentioned in Section 25.7 (see Equation (25.5)), the risk-weighted assets (RWA) are calculated as 12.5 times the capital required

$$\text{RWA} = 12.5 \times \text{EAD} \times \text{LGD} \times (\text{WCDR} - \text{PD}) \times \text{MA}$$

so that the capital is 8% of RWA, 4% of which must be Tier 1.

Under the Foundation IRB approach, banks supply PD while LGD, EAD, and M are supervisory values set by the Basel Committee. PD is subject to a floor of 0.03% for bank and corporate exposures. LGD is set at 45% for senior claims and 75% for subordinated claims. When there is eligible collateral, in order to correspond to the comprehensive approach that we described earlier, LGD is reduced by the ratio of the adjusted value of the collateral to the adjusted value of the exposure, both calculated using the comprehensive approach. For derivatives, the EAD is calculated using the CEM “current exposure plus add-on” approach of Basel I and includes the impact of netting. M is set at 2.5 in most circumstances.

Under the Advanced IRB approach, banks supply their own estimates of the PD, LGD, EAD, and M for corporate, sovereign, and bank exposures. The PD can be reduced by credit mitigants such as credit triggers. (As in the case of the Foundation IRB approach, it is subject to a floor of 0.03% for bank and corporate exposures.) The two main factors influencing the LGD are the seniority of the debt and the collateral.

In calculating EAD, banks can with regulatory approval use their own models. As mentioned, in the case of derivatives, the model is likely to involve Monte Carlo simulation to determine how expected exposure (after netting and collateral) will vary over the next year.

The capital given by equation (25.9) is intended to be sufficient to cover unexpected losses over a one-year period that we are 99.9% certain will not be exceeded. (As discussed earlier, the expected losses should be covered by a bank in the way it prices its products.) The WCDR is the default rate that (theoretically) happens once every thousand years. The Basel Committee reserved the right to apply a scaling factor (less than or greater than 1.0) to the result of the calculations in equation (25.9) if it finds that the aggregate capital requirements are too high or low. A typical scaling factor is 1.06.

Example 25.7

Suppose that the assets of a bank consist of \$100 million of loans to A-rated corporations. The PD for the corporations is estimated as 0.1% and the LGD is 60%. The average maturity is 2.5 years for the corporate loans. This means that

$$b = [0.11852 - 0.05478 \times \ln(0.001)]^2 = 0.247$$

so that

$$MA = \frac{1}{1 - 1.5 \times 0.247} = 1.59$$

From Table 25.6, the WCDR is 3.4%. Under the Basel II IRB approach, the risk-weighted assets for the corporate loans are

$$12.5 \times 100 \times 0.6 \times (0.034 - 0.001) \times 1.59 = 39.3$$

or \$39.3 million. This compares with \$100 million under Basel I and \$50 million under the standardized approach of Basel II. (See Examples 25.1 and 25.5 where a \$100 million corporate loan is part of the portfolio.)

25.8.6 Retail Exposures

The model underlying the calculation of capital for retail exposures in Basel II is similar to that underlying the calculation of corporate, sovereign, and banking exposures. However, the Foundation IRB and Advanced IRB approaches are merged and all banks using the IRB approach provide their own estimates of PD, EAD, and LGD. There is no maturity adjustment, MA. The capital requirement is therefore

$$EAD \times LGD \times (WCDR - PD)$$

and the risk-weighted assets are

$$12.5 \times EAD \times LGD \times (WCDR - PD)$$

Table 25.7 Relationship between One-Year 99.9% WCDR and PD for Retail Exposures

PD	0.1%	0.5%	1.0%	1.5%	2.0%
WCDR	2.1%	6.3%	9.1%	11.0%	12.3%

WCDR is calculated as in equation (25.6). For residential mortgages, ρ is set equal to 0.15 in this equation.¹⁸ For qualifying revolving exposures, ρ is set equal to 0.04. For all other retail exposures, a relationship between ρ and PD is specified for the calculation of WCDR. This is

$$\rho = 0.03 \frac{1 - \exp(-35 \times \text{PD})}{1 - \exp(-35)} + 0.16 \left[1 - \frac{1 - \exp(-35 \times \text{PD})}{1 - \exp(-35)} \right]$$

Because $\exp(-35)$ is a very small number, this formula is to all intents and purposes

$$\rho = 0.03 + 0.13e^{-35 \times \text{PD}} \quad (25.11)$$

Comparing equation (25.11) with equation (25.8), we see that correlations are assumed to be much lower for retail exposures than for corporate exposures. Table 25.7 is the table corresponding to Table 25.6 for retail exposures.

Example 25.8

Suppose that the assets of a bank consist of \$50 million of residential mortgages where the PD is 0.005 and the LGD is 20%. In this case, $\rho = 0.15$ and

$$\text{WCDR} = N \left[\frac{N^{-1}(0.005) + \sqrt{0.15} N^{-1}(0.999)}{\sqrt{1 - 0.15}} \right] = 0.067$$

The risk-weighted assets are

$$12.5 \times 50 \times 0.2 \times (0.067 - 0.005) = 7.8$$

or \$7.8 million. This compares with \$25 million under Basel I and \$17.5 million under the Standardized Approach of Basel II. (See Examples 25.1 and 25.5 where \$50 million of residential mortgages is part of the portfolio.)

¹⁸In the light of experience during the crisis, this is probably too low.

25.8.7 *Guarantees and Credit Derivatives*

The approach traditionally taken by the Basel Committee for handling guarantees and credit derivatives such as credit default swaps is the credit substitution approach. Suppose that a AA-rated company guarantees a loan to a BBB-rated company. For the purposes of calculating capital, the credit rating of the guarantor is substituted for the credit rating of the borrower so that capital is calculated as though the loan had been made to the AA-rated company. This overstates the credit risk because, for the lender to lose money, both the guarantor and the borrower must default (with the guarantor defaulting before the borrower).¹⁹ The Basel Committee has addressed this issue. In July 2005, it published a document concerned with the treatment of double defaults under Basel II.²⁰ As an alternative to using the credit substitution approach, the capital requirement can be calculated as the capital that would be required without the guarantee multiplied by $0.15 + 160 \times PD_g$ where PD_g is the one-year probability of default of the guarantor.

25.9 Operational Risk Capital Under Basel II

In addition to changing the way banks calculate credit risk capital, Basel II requires banks to keep capital for operational risk. This is the risk of losses from situations where the bank's procedures fail to work as they are supposed to or where there is an adverse external event such as a fire in a key facility.

There were three approaches to calculating capital for operational risk in Basel II:

1. The Basic Indicator Approach
2. The Standardized Approach
3. The Advanced Measurement Approach

Which of these was used depended on the sophistication of the bank. The simplest approach is the Basic Indicator Approach. This sets the operational risk capital equal to the bank's average annual gross income over the last three years multiplied by 0.15.²¹ The Standardized Approach is similar to the basic indicator approach except that a different factor is applied to the gross income from different business lines. In the Advanced Measurement Approach, the bank uses its own internal models to calculate the operational risk loss that it is 99.9% certain will not be exceeded in one year. Similarly to the way credit risk capital is calculated in the IRB approach, operational risk capital

¹⁹Credit default swaps, which we discussed in Chapter 17, provide a type of insurance against default and are handled similarly to guarantees for regulatory purposes.

²⁰See Bank for International Settlements, "The Application of Basel II to Trading Activities and the Treatment of Double Defaults," July 2005, available on www.bis.org.

²¹Gross income defined as net interest income plus non-interest income. Net interest income is the excess of income earned on loans over interest paid on deposits and other instruments that are used to fund the loans. Years where gross income is negative are not included in the calculations.

was set equal to this loss minus the expected loss. One advantage of the advanced measurement approach was that it allowed banks to recognize the risk-mitigating impact of insurance contracts subject to certain conditions. However, because there was very little consistency between the calculations of different banks, it has now been replaced by the standardized measurement approach described in Section 20.4.

25.10 Pillar 2: Supervisory Review

Pillar 2 of Basel II is concerned with the supervisory review process. Four key principles of supervisory review were specified:

1. Banks should have a process for assessing their overall capital adequacy in relation to their risk profile and a strategy for maintaining their capital levels.
2. Supervisors should review and evaluate banks' internal capital adequacy assessments and strategies, as well as their ability to monitor and ensure compliance with regulatory capital ratios. Supervisors should take appropriate supervisory action if they are not satisfied with the result of this process.
3. Supervisors should expect banks to operate above the minimum regulatory capital and should have the ability to require banks to hold capital in excess of this minimum.
4. Supervisors should seek to intervene at an early stage to prevent capital from falling below the minimum levels required to support the risk characteristics of a particular bank and should require rapid remedial action if capital is not maintained or restored.

The Basel Committee suggested that regulators pay particular attention to interest rate risk in the banking book, credit risk, and operational risk. Key issues in credit risk are stress tests used, default definitions used, credit risk concentration, and the risks associated with the use of collateral, guarantees, and credit derivatives.

The Basel Committee also stressed that there should be transparency and accountability in the procedures used by bank supervisors. This is particularly important when a supervisor exercises discretion in the procedures used or sets capital requirements above the minimum specified in Basel II.

25.11 Pillar 3: Market Discipline

Pillar 3 of Basel II is concerned with increasing the disclosure by a bank of its risk assessment procedures and capital adequacy. The extent to which regulators can force banks to increase their disclosure varies from jurisdiction to jurisdiction. However, banks are unlikely to ignore directives on this from their supervisors, given the potential of supervisors to make their life difficult. Also, in some instances, banks have to increase their disclosure in order to be allowed to use particular methodologies for calculating capital.

Regulatory disclosures are likely to be different in form from accounting disclosures and need not be made in annual reports. It is largely left to the bank to choose disclosures that are material and relevant. Among the items that banks should disclose are:

1. The entities in the banking group to which Basel II is applied and adjustments made for entities to which it is not applied
2. The terms and conditions of the main features of all capital instruments
3. A list of the instruments constituting Tier 1 capital and the amount of capital provided by each item
4. The total amount of Tier 2 capital
5. Capital requirements for credit, market, and operational risk
6. Other general information on the risks to which a bank is exposed and the assessment methods used by the bank for different categories of risk
7. The structure of the risk management function and how it operates

25.12 Solvency II

As discussed in Section 3.11, there are no international standards for the regulation of insurance companies. In the United States, insurance companies are regulated at the state level with some input from the Federal Insurance Office and the National Association of Insurance Commissioners. In Europe, the regulation of insurance companies is handled by the European Union. The long-standing regulatory framework in the European Union, known as Solvency I, was replaced by Solvency II in 2016. Whereas Solvency I calculates capital only for underwriting risks, Solvency II considers investment risks and operational risks as well.

Solvency II has many similarities to Basel II. There are three pillars. Pillar 1 is concerned with the calculation of capital requirements and the types of capital that are eligible. Pillar 2 is concerned with the supervisory review process. Pillar 3 is concerned with the disclosure of risk management information to the market. The three pillars are therefore analogous to the three pillars of Basel II.

Pillar 1 of Solvency II specifies a *minimum capital requirement* (MCR) and a *solvency capital requirement* (SCR). If its capital falls below the SCR level, an insurance company should, at minimum, deliver to the supervisor a plan to restore capital to above the SCR level. The supervisor might require the insurance company to take particular measures to correct the situation. The MCR is regarded as an absolute minimum level of capital. If capital drops below the MCR level, supervisors may assume control of the insurance company and prevent it from taking new business. It might force the insurance company into liquidation, transferring its policies to another company. The MCR is typically between 25% and 45% of the SCR.

There are two ways to calculate the SCR: the standardized approach and the internal models approach. The internal models approach involves a VaR calculation with a one-year time horizon and a 99.5% confidence limit. (The confidence level is therefore less than the 99.9% confidence level used in Pillar 1 of Basel II.) Longer time horizons

with lower confidence levels are also allowed when the protection provided is considered equivalent. The SCR involves a capital charge for investment risk, underwriting risk, and operational risk. Investment risk is subdivided into market risk and credit risk. Underwriting risk is subdivided into risk arising from life insurance, non-life insurance (i.e., property and casualty), and health insurance.

Capital should be adequate to deal with large adverse events. Examples of the events considered in quantitative impact studies are:

1. A 32% decrease in global stock markets
2. A 20% decrease in real estate prices
3. A 20% change in foreign exchange rates
4. Specified catastrophic risk scenarios affecting property and casualty payouts
5. Health care costs increasing by a factor times the historical standard deviation of costs
6. A 10% increase in mortality rates
7. A 25% decrease in mortality rates
8. A 10% increase in expenses

The internal models are required to satisfy three tests. The first is a statistical quality test. This is a test of the soundness of the data and methodology used in calculating VaR. The second is a calibration test. This is a test of whether risks have been assessed in accordance with a common SCR target criterion. The third is a use test. This is a test of whether the model is genuinely relevant to and used by risk managers.

There are three types of capital in Solvency II. Tier 1 capital consists of equity capital, retained earnings, and other equivalent funding sources. Tier 2 capital consists of liabilities that are subordinated to policyholders and satisfy certain criteria concerning their availability in wind-down scenarios. Tier 3 capital consists of liabilities that are subordinated to policyholders and do not satisfy these criteria. Similarly to Basel II, the amount of capital that must be Tier 1, Tier 1 plus Tier 2, and Tier 1 plus Tier 2 plus Tier 3 is specified.

Summary

This chapter has provided an overview of capital requirements for banks up to 2007. The way in which regulators calculate the minimum capital a bank is required to hold has changed dramatically since the 1980s. Prior to 1988, regulators determined capital requirements by specifying minimum ratios for capital to assets or maximum ratios for assets to capital. In the late 1980s, both bank supervisors and the banks themselves agreed that changes were necessary. Derivatives trading was increasing fast. Also, banks were competing globally and it was considered important to create a level playing field by making regulations uniform throughout the world.

The 1988 Basel Accord assigned capital for credit risk for both on- and off-balance-sheet exposures. This involved calculating a risk-weighted asset for each item. The risk-weighted assets for an on-balance-sheet loan were calculated by

multiplying the principal by a risk weight for the counterparty. In the case of derivatives such as swaps, banks were first required to calculate credit equivalent amounts. The risk-weighted assets were obtained by multiplying the credit equivalent amount by a risk weight for the counterparty. Banks were required to keep capital equal to 8% of the total risk-weighted assets. In 1995, the capital requirements for credit risk were modified to incorporate netting. As a result of an amendment introduced in 1996, the Accord was changed to include a capital charge for market risk. Sophisticated banks could base the capital charge on a value-at-risk calculation.

Basel II was proposed in 1999 and implemented by many banks in about 2007. It led to no immediate change to the capital requirement for market risk. Credit risk capital was calculated in a more sophisticated way than previously to reflect either (a) the credit ratings of obligors or (b) estimates made by the bank in conjunction with a default correlation parameter specified by regulators. In addition, there was a capital charge for operational risk.

Solvency II is a regulatory framework for insurance companies that was implemented by the European Union in 2016. It prescribes minimum capital levels for investment risk, underwriting risk, and operational risk. The structure of Solvency II is broadly similar to Basel II.

Further Reading

- Bank for International Settlements. "Basel II: International Convergence of Capital Measurement and Capital Standards." June 2006, www.bis.org.
- Gordy, M. B. "A Risk-Factor Model Foundation for Ratings-Based Bank Capital Ratios." *Journal of Financial Intermediation* 12 (2003): 199–232.
- Lopez, J. A. "The Empirical Relationship between Average Asset Correlation, Firm Probability of Default, and Asset Size." *Journal of Financial Intermediation* 13, no. 2 (2004): 265–283.
- Vasicek, O. "Probability of Loss on a Loan Portfolio." Working Paper, KMV, 1987. (Published in *Risk* in December 2002 under the title "Loan Portfolio Value.")

Practice Questions and Problems (Answers at End of Book)

- 25.1 "When a steel company goes bankrupt, other companies in the same industry benefit because they have one less competitor. But when a bank goes bankrupt other banks do not necessarily benefit." Explain this statement.
- 25.2 "The existence of deposit insurance makes it particularly important for there to be regulations on the amount of capital banks hold." Explain this statement.
- 25.3 An interest rate swap involves the exchange of a fixed rate of interest for a floating rate of interest with both being applied to the same principal. The principals are not exchanged. What is the nature of the credit risk for a bank when it enters into a five-year interest rate swap with a notional principal of \$100 million? Assume the swap is worth zero initially.

- 25.4 In a currency swap, interest on a principal in one currency is exchanged for interest on a principal in another currency. The principals in the two currencies are exchanged at the end of the life of the swap. Why is the credit risk on a currency swap greater than that on an interest rate swap?
- 25.5 A four-year interest rate swap currently has a negative value to a financial institution. Is the financial institution exposed to credit risk on the transaction? Explain your answer. How would the capital requirement be calculated under Basel I?
- 25.6 Estimate the capital required under Basel I for a bank that has the following transactions with a corporation. Assume no netting.
- (a) A nine-year interest rate swap with a notional principal of \$250 million and a current market value of $-\$2$ million.
 - (b) A four-year interest rate swap with a notional principal of \$100 million and a current value of $\$3.5$ million.
 - (c) A six-month derivative on a commodity with a principal of \$50 million that is currently worth \$1 million.
- 25.7 What is the capital required in Problem 25.6 under Basel I assuming that the 1995 netting amendment applies?
- 25.8 All the derivatives transactions a bank has with a corporate client have a positive value to the bank. What is the value to the bank of netting provisions in its master agreement with the client?
- 25.9 Explain why the final stage in the Basel II calculations for credit risk (IRB), market risk, and operational risk is to multiply by 12.5.
- 25.10 What is the difference between the trading book and the banking book for a bank? A bank currently has a loan of \$10 million to a corporate client. At the end of the life of the loan, the client would like to sell debt securities to the bank instead of borrowing. How does this potentially affect the nature of the bank's regulatory capital calculations?
- 25.11 Under Basel I, banks do not like lending to highly creditworthy companies and prefer to help them issue debt securities. Why is this? Do you think this changed as a result of Basel II?
- 25.12 Banks sometimes use what is referred to as regulatory arbitrage to reduce their capital. What do you think this means?
- 25.13 Equation (25.9) gives the formula for the capital required under Basel II. It involves four terms being multiplied together. Explain each of these terms.
- 25.14 Explain the difference between the simple and the comprehensive approach for adjusting capital requirements for collateral.
- 25.15 Explain the difference between the standardized approach, the IRB approach, and the Advanced IRB approach for calculating credit risk capital under Basel II.
- 25.16 Explain the difference between the basic indicator approach, the standardized approach, and the advanced measurement approach for calculating operational risk capital under Basel II.
- 25.17 Suppose that the assets of a bank consist of \$200 million of retail loans (not mortgages). The PD is 1% and the LGD is 70%. What are the risk-weighted assets

under the Basel II IRB approach? What are the Tier 1 and Tier 2 capital requirements?

- 25.18 Section 11.10 discusses how statistics can be used to accept or reject a VaR model. Section 25.6 discusses guidelines for bank supervisors in setting the VaR multiplier m_c . It explains that, if the number of exceptions in 250 trials is five or more, then m_c is increased. What is the chance of five or more exceptions if the VaR model is working well?

Further Questions

- 25.19 Why is there an add-on amount in Basel I for derivatives transactions? “Basel I could be improved if the add-on amount for a derivatives transaction depended on the value of the transaction.” How would you argue this viewpoint?
- 25.20 Estimate the capital required under Basel I for a bank that has the following transactions with another bank. Assume no netting.
- (a) A two-year forward contract on a foreign currency, currently worth \$2 million, to buy foreign currency worth \$50 million
 - (b) A long position in a six-month option on the S&P 500. The principal is \$20 million and the current value is \$4 million.
 - (c) A two-year swap involving oil. The principal is \$30 million and the current value of the swap is $-\$5$ million.

What difference does it make if the netting amendment applies?

- 25.21 A bank has the following transaction with a AA-rated corporation
- (a) A two-year interest rate swap with a principal of \$100 million that is worth \$3 million
 - (b) A nine-month foreign exchange forward contract with a principal of \$150 million that is worth $-\$5$ million
 - (c) A long position in a six-month option on gold with a principal of \$50 million that is worth \$7 million

What is the capital requirement under Basel I if there is no netting? What difference does it make if the netting amendment applies? What is the capital required under Basel II when the standardized approach is used?

- 25.22 Suppose that the assets of a bank consist of \$500 million of loans to BBB-rated corporations. The PD for the corporations is estimated as 0.3%. The average maturity is three years and the LGD is 60%. What is the total risk-weighted assets for credit risk under the Basel II IRB approach? How much Tier 1 and Tier 2 capital is required? How does this compare with the capital required under the Basel II standardized approach and under Basel I?

Chapter 26

Basel II.5, Basel III, and Other Post-Crisis Changes

It was perhaps unfortunate for Basel II that its implementation date coincided, at least approximately, with the start of the worst crisis that financial markets had experienced since the 1930s. Some commentators have blamed Basel II for the Global Financial Crisis. They point out that it was a move toward self-regulation where banks, when calculating regulatory capital, had the freedom to use their own estimates of model inputs such as PD, LGD, and EAD. However, blaming Basel II may be unfair because, as explained in Chapter 7, the seeds of the crisis were sown well before Basel II was implemented.¹

This chapter starts by discussing what has become known as Basel II.5. This is a collection of changes to the calculation of market risk capital that was put in place by the Basel Committee on Banking Supervision following the large losses experienced by banks during the crisis. The implementation date for Basel II.5 was December 31, 2011.

The chapter then moves on to consider Basel III, which was a much bigger post-crisis overhaul of bank regulations. A first set of regulations were published in December 2010 and the Basel II regulations were finalized in December 2017. (New rules for market risk, which are part of Basel III but sometimes referred to as Basel IV, are discussed in Chapter 27.) The implementation of Basel III has been phased in over several years.

¹Also, the United States was behind other countries in implementing Basel II. If Basel II had been fully implemented by the start of the crisis, capital levels at U.S. banks would probably have been lower.

The chapter also discusses other regulations that have been introduced since the 2008 crisis to complement the work of the Basel Committee. These include the Dodd–Frank Act, which was signed into law by President Barack Obama in the United States on July 21, 2010, and rules introduced in the European Union and the United Kingdom.

26.1 Basel II.5

During the credit crisis, it was recognized that some changes were necessary to the calculation of capital for market risk. These changes are referred to as Basel II.5 and, as already mentioned, the implementation date for them was December 31, 2011.² There are three changes involving:

1. The calculation of a stressed VaR;
2. A new incremental risk charge; and
3. A comprehensive risk measure for instruments dependent on credit correlation.

The measures have the effect of greatly increasing the market risk capital that large banks are required to hold.

26.1.1 *Stressed VaR*

The 1996 Amendment to Basel I, where capital was first required for market risk, allowed banks to base capital on a 10-day 99% VaR measure. Most banks use historical simulation to calculate VaR. This is described in Chapter 12. The assumption underlying historical simulation was that the changes in market variables during the following day would be a random sample from their daily changes (percentage or actual) observed during the previous one to four years. The 2003–2006 period was one where the volatility of most market variables was low. As a result, the market risk VaRs calculated during this period for regulatory capital purposes were also low. Furthermore, the VaRs continued to be too low for a period of time after the onset of the crisis, because much of the data used to calculate them continued to come from a low-volatility period.

This led the Basel Committee to introduce the stressed VaR measure. As explained in Section 12.1.3, stressed VaR is determined by basing calculations on how market variables moved during a 250-day (12-month) period of stressed market conditions, rather than on how they moved during the past one to four years. The historical simulation calculations to arrive at a stressed VaR measure assume that the changes in market variables during the next day are a random sample from their daily changes observed during the 250-day period of stressed market conditions.

Basel II.5 requires banks to calculate two VaRs. One is the usual VaR (based on the previous one to four years of market movements). The other is the stressed VaR

²See Basel Committee on Banking Supervision, “Revisions to the Basel II Market Risk Framework,” February 2011.

(calculated from a stressed period of 250 days). The two VaR measures are combined to calculate a total capital charge. The formula for the total capital charge is

$$\max(\text{VaR}_{t-1}, m_c \times \text{VaR}_{\text{avg}}) + \max(\text{sVaR}_{t-1}, m_s \times \text{sVaR}_{\text{avg}})$$

where VaR_{t-1} and sVaR_{t-1} are the VaR and stressed VaR (with a 10-day time horizon and a 99% confidence level) calculated on the previous day. The variables VaR_{avg} and sVaR_{avg} are the average of VaR and stressed VaR (again with a 10-day time horizon and a 99% confidence level) calculated over the previous 60 days. The parameters m_s and m_c are multiplicative factors that are determined by bank supervisors and are at minimum equal to three. As explained in Section 25.6, the market risk capital requirement prior to Basel II.5 was

$$\max(\text{VaR}_{t-1}, m_c \times \text{VaR}_{\text{avg}})$$

Because stressed VaR is always at least as great as VaR, the formula shows that (assuming $m_c = m_s$) the impact of this new rule is to at least double the capital requirement. In practice, a trebling of the capital requirement as a result of the change was not unusual.

Originally it was considered that 2008 would constitute a good one-year period for the calculation of stressed VaR. Later it was realized that the one-year period chosen should reflect a bank's portfolio. A bank is therefore required to search for a one-year period that would be particularly stressful for its current portfolio. The stressed period used by one bank is therefore not necessarily the same as that used by another bank.

26.1.2 Incremental Risk Charge

In 2005, the Basel Committee became concerned that exposures in the trading book were attracting less capital than similar exposures in the banking book. Consider a bond. If held in the trading book, the capital would be calculated by applying a multiplier to the 10-day 99% VaR, as discussed in Section 25.6. If held in the banking book (and treated like a loan), capital for the bond would be calculated using VaR with a one-year time horizon and a 99.9% confidence level, as discussed in Section 25.8. The trading-book calculation usually gave rise to a much lower capital charge than the banking-book calculation. As a result, banks tended whenever possible to hold credit-dependent instruments in the trading book.³

Regulators proposed what is referred to as an incremental default risk charge (IDRC) in 2005 for instruments in the trading book that were sensitive to default risk. The effect of this was that the capital requirement for these instruments equaled the maximum of the capital using trading book calculations and the capital using banking book calculations. In 2008, the Basel Committee recognized that most of the losses in the credit market

³If a bank created ABSs from loans in the banking book, as described in Chapter 7, and then bought all the resultant tranches for its trading book, regulatory capital requirements would be lowered even though the bank's risks would be unchanged. This was one reason why banks wanted to securitize loans in the banking book.

turmoil of 2007 and 2008 were from changes in credit ratings, widening of credit spreads, and loss of liquidity, rather than solely as a result of defaults. It therefore amended its previous proposals to reflect this. The new requirement became known as the incremental risk charge (IRC).⁴

The IRC requires banks to calculate a one-year 99.9% VaR for losses from credit-sensitive products in the trading book taking both credit rating changes and defaults into account. Like the IDRC, the aim was to set capital equal to the maximum of that obtained using trading book calculations and that obtained using banking book calculations. Because the instruments subject to the IRC are in the trading book, it is assumed that a bank will have the opportunity to rebalance its portfolio during the course of the year so that default risk is mitigated. Banks are therefore required to estimate a liquidity horizon for each instrument subject to the IRC. The liquidity horizon represents the time required to sell the position or to hedge all material risks in a stressed market.

Suppose that the liquidity horizon for a bond with a credit rating of A is three months. For the purposes of the calculation of VaR over a one-year time horizon, the bank assumes that at the end of three months, if the bond's rating has changed or if it has defaulted, it is replaced by an A-rated bond similar to that held at the beginning of the period. The same thing happens at the end of six months and at the end of nine months. This is known as the *constant level of risk* assumption.

The impact of the constant level of risk assumption is that it is less likely that there will be a default. Instead, small losses are realized from ratings downgrades when rebalancing takes place. The assumption typically has the effect of reducing the one-year 99.9% VaR.⁵ The minimum liquidity horizon for IRC is specified by the Basel Committee as three months.

The IRC therefore provides a measure of the default and credit migration risks of credit products over a one-year horizon at a 99.9% confidence level, taking into account the liquidity horizons of individual positions or sets of positions.

26.1.3 The Comprehensive Risk Measure

The comprehensive risk measure (CRM) is designed to take account of risks in what is known as the correlation book. This is the portfolio of instruments such as asset-backed securities (ABSs) and collateralized debt obligations (CDOs) that are sensitive to the correlation between the default risks of different assets. These instruments were discussed in Chapter 7. Suppose a bank has a AAA-rated tranche of an ABS. In the normal market environment, there is very little risk of losses from the tranche. However, in stressed market environments when correlations increase, the tranche is vulnerable—as became apparent during the 2007–2009 crisis.

⁴Basel Committee on Banking Supervision, “Guidelines for Computing Capital for Incremental Risk in the Trading Book,” July 2009.

⁵See C. Finger, “CreditMetrics and Constant Level of Risk,” MSCI, 2010, for a discussion of the constant level of risk assumption.

Table 26.1 Standardized Capital Charge for Correlation-Dependent Instruments

External Credit Assessment	AAA to AA–	A+ to A–	BBB+ to BBB–	BB+ to BB–	Below BB– or Unrated
Securitizations	1.6%	4%	8%	28%	Deduction
Resecuritizations	3.2%	8%	18%	52%	Deduction

The CRM is a single capital charge replacing the incremental risk charge and the specific risk charge for instruments dependent on credit correlation. The Basel II.5 standardized approach for calculating the CRM is summarized in Table 26.1. Given the experience of the securitization market during the crisis (see Chapter 7), it is not surprising that capital charges are higher for resecuritizations (e.g., ABS CDOs) than for securitizations (e.g., ABSs). A deduction means that the principal amount is subtracted from capital, which is equivalent to a 100% capital charge.

Basel II.5 allows banks, with supervisory approval, to use their internal models to calculate the CRM. The models developed by banks have to be quite sophisticated to be approved by bank supervisors. For example, they must capture the cumulative impact of multiple defaults, credit spread risk, the volatility of implied correlations, the relationship between credit spreads and implied correlations, recovery rate volatility, the risk of hedge slippage, and potential hedge rebalancing costs. A routine and rigorous program of stress testing is also required. The capital charge calculated from an internal model is subject to a floor calculated from the capital given by the standardized approach.

26.2 Basel III

Following the 2007–2009 Global Financial Crisis, the Basel Committee realized that a major overhaul of Basel II was necessary. Basel II.5 increased capital requirements for market risk. The Basel Committee wanted to increase equity capital requirements further. In addition, it considered that the definition of equity capital needed to be tightened and that regulations were needed to address liquidity risk.

Basel III proposals were first published in December 2009 soon after the end of the Global Financial Crisis. Following comments from banks, a quantitative impact study, and a number of international summits, a first set of regulations was published in December 2010.⁶ They were finalized in 2017.

There are six parts to the regulations:

1. Capital Definition and Requirements
2. Capital Conservation Buffer

⁶See Basel Committee on Banking Supervision, “Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems,” June 2011, and Basel Committee on Banking Supervision, “Basel III: International Framework for Liquidity Risk Measurement Standards and Monitoring,” December 2010.

3. Countercyclical Buffer
4. Leverage Ratio
5. Liquidity Risk
6. Counterparty Credit Risk

The regulations were originally scheduled to be introduced between 2013 and 2015. However, for various reasons implementation of many of the Basel III requirements has been delayed to January 2023.

26.2.1 Capital Definition and Requirements

Under Basel III, a bank's total capital consists of:

1. Tier 1 equity capital
2. Additional Tier 1 capital
3. Tier 2 capital

There is no Tier 3 capital.

Tier 1 equity capital (also referred to as common equity Tier 1 or CET1 capital) includes share capital and retained earnings but does not include goodwill or deferred tax assets. It must be adjusted downward to reflect defined benefit pension plan deficits but is not adjusted upward to reflect defined benefit plan surpluses. (See Section 3.12 for a discussion of defined benefit plans.) Changes in retained earnings arising from securitized transactions are not counted as part of capital for regulatory purposes. The same is true of changes in retained earnings arising from the bank's own credit risk. (The latter is referred to as DVA and is discussed in Chapter 18.) There are rules relating to the inclusion of minority interests and capital issued by consolidated subsidiaries.

The additional Tier 1 (AT1) capital category consists of items, such as non-cumulative preferred stock, that were previously Tier 1 but are not common equity. Tier 2 capital includes debt that is subordinated to depositors with an original maturity of five years.

Common equity is referred to by the Basel Committee as "going-concern capital." When the bank is a going concern (i.e., has positive equity capital), common equity absorbs losses. Tier 2 capital is referred to as "gone-concern capital." When the bank is no longer a going concern (i.e., has negative capital), losses have to be absorbed by Tier 2 capital. Tier 2 capital ranks below depositors in a liquidation. While Tier 2 capital remains positive, depositors should in theory be repaid in full.

The capital requirements are as follows:

1. Tier 1 equity capital must be at least 4.5% of risk-weighted assets at all times.
2. Total Tier 1 capital (Tier 1 equity capital plus additional Tier 1 capital) must be at least 6% of risk-weighted assets at all times.
3. Total capital (total Tier 1 plus Tier 2) must be at least 8% of risk-weighted assets at all times.

Table 26.2 Dividend Restrictions Arising from the Capital Conservation Buffer

Tier 1 Equity Capital Ratio	Minimum Percent of Earnings Retained
4.000% to 5.125%	100%
5.125% to 5.750%	80%
5.750% to 6.375%	60%
6.375% to 7.000%	40%
>7%	0%

Basel I required Tier 1 equity capital to be at least 2% of risk-weighted assets and total Tier 1 capital to be at least 4% of risk-weighted assets. The Basel III rules are much more demanding because (a) these percentages have been increased and (b) the definition of what qualifies as equity capital for regulatory purposes has been tightened. However, the Tier 1 plus Tier 2 requirement is the same as under Basel I and Basel II.

The Basel Committee also calls for more capital for “systemically important” banks, as will be discussed later in this chapter.

26.2.2 Capital Conservation Buffer

In addition to the capital requirements just mentioned, Basel III requires a capital conservation buffer in normal times consisting of a further amount of CET1 equal to 2.5% of risk-weighted assets. This provision is designed to ensure that banks build up capital during normal times so that it can be run down when losses are incurred during periods of financial difficulties. (The argument in favor of this is that it is much easier for banks to raise capital during normal times than during periods of stressed market conditions.) In circumstances where the capital conservation buffer has been wholly or partially used up, banks are required to constrain their dividends until the capital has been replenished. The dividend rules are summarized in Table 26.2. For example, if CET1 is 5.5% of risk-weighted assets, the minimum retained earnings is 80% so that the maximum dividends as a percent of retained earnings is 20%. The difference between the 4.5% basic equity capital requirement and the 7.0% equity capital requirement that includes the capital conservation buffer is in some ways analogous to the difference between the MCR and SCR in Solvency II (see Section 25.12).

The capital conservation buffer means that the Tier 1 equity capital that banks are required to keep in normal times (excluding other requirements that will be discussed later) is 7% of risk-weighted assets; total Tier 1 capital is required to be at least 8.5% of risk-weighted assets; Tier 1 plus Tier 2 capital is required to be at least 10.5% of risk-weighted assets. These numbers can decline to 4.5%, 6%, and 8% in stressed market conditions (because of losses), but banks are then under pressure to bring capital back up to the required levels. One of the consequences of the increased equity capital requirement is that banks may find it difficult to achieve the returns on equity that they had

Table 26.3 Dividend Restrictions Arising from the Capital Conservation Buffer and 2.5% Countercyclical Buffer

Tier 1 Equity Capital Ratio	Minimum Percent of Earnings Retained
4.50% to 5.75%	100%
5.75% to 7.00%	80%
7.00% to 8.25%	60%
8.25% to 9.50%	40%
>9.50%	0%

during the 1990 to 2006 period. However, bank shareholders can console themselves that bank stock is less risky as a result of the extra capital.

26.2.3 Countercyclical Buffer

In addition to the capital conservation buffer, Basel III has specified a countercyclical buffer. This is similar to the capital conservation buffer, but the extent to which it is implemented in a particular country is left to the discretion of national authorities. The buffer is intended to provide protection for the cyclicity of bank earnings. The buffer can be set to between 0% and 2.5% of total risk-weighted assets and must be met with Tier 1 equity capital (CET1).

For jurisdictions where the countercyclical buffer is non-zero, Table 26.2 is modified. For example, when the countercyclical buffer is set at its maximum level of 2.5%, it is replaced by Table 26.3.

26.2.4 Leverage Ratio

In addition to the capital requirements based on risk-weighted assets, Basel III specifies a minimum leverage ratio of 3%.⁷ The leverage ratio is the ratio of a capital measure to an exposure measure. The capital measure is total Tier 1 capital. The exposure measure is the sum of (a) on-balance-sheet exposures, (b) derivatives exposures, (c) securities financing transaction exposures, and (d) off-balance-sheet items. No risk-weighting adjustments are made. The on-balance-sheet exposures include all assets on the balance sheet. Derivatives exposures are calculated as “replacement cost plus add-on” similar to the way they are calculated under Basel I (see Section 25.3). Securities financing transaction exposures include transactions such as repurchase agreements and security lending/borrowing when the transaction does not lead to balance sheet assets. Off-balance-sheet items include loan commitments, loan substitutes, acceptances, and letters of credit.

Regulators in some countries have indicated that they regard Basel’s leverage ratio of 3% as too low. In April 2014, regulators in the United States proposed rules requiring

⁷See Bank for International Settlements, “Basel III Leverage Ratio Framework and Disclosure Requirements,” January 2014.

a 5% leverage ratio for eight large bank holding companies (which, as we will explain later, are known as G-SIBs), and 6% for the FDIC-insured subsidiaries of these bank holding companies. In October 2014, the Bank of England's Financial Policy Committee increased the leverage ratio for UK banks to 4.05%, a level that could be increased to 4.95% in boom times to rein in excessive lending. In China the leverage ratio has been set at 4%.

Why did the Basel Committee introduce the leverage ratio? The reason is that regulators thought that banks had too much discretion in the way risk-weighted assets were calculated. They have far less discretion in the way “total exposure” is calculated. It should be emphasized that this does not mean that regulators are discarding capital requirements based on risk-weighted assets. They require banks to satisfy both (a) the ratios of capital to risk-weighted assets mentioned earlier in this chapter and (b) the ratio of capital to non-risk-weighted exposure leverage requirement explained here.

One of the two capital ratios is likely to be the critical ratio for a bank (i.e., the ratio it is closest to not complying with). If the risk-weighted assets ratio is the critical one, then arguably the leverage ratio can be justified because it provides useful additional information to regulators. But if the leverage ratio is the critical one (and there is some evidence that at times it has been for some U.S. banks), a bank might be encouraged to hold risky assets because they have the same effect on the leverage ratio as safe assets but provide a higher expected return. This could be an unintended adverse consequence for regulators.

26.2.5 Liquidity Risk

Prior to the financial crisis, the focus of the Basel regulations had been on ensuring that banks had sufficient capital for the risks they were taking. It turned out that many of the problems encountered by financial institutions during the crisis were not as a result of a shortage of capital. They were instead a result of liquidity risks taken by the banks.

Liquidity risks arise because there is a tendency for banks to finance long-term needs with short-term funding, such as commercial paper. Provided the bank is perceived by the market to be financially healthy, this is usually not a problem.⁸ Suppose that a bank uses 90-day commercial paper to fund its activities. When one 90-day issue of commercial paper matures, the bank refinances with a new issue; when the new issue matures, it refinances with another issue; and so on. However, as soon as the bank experiences financial difficulties—or is thought to be experiencing financial difficulties—it is liable to become impossible for the bank to roll over its commercial paper. This type of problem led to the demise of Northern Rock in the United Kingdom and Lehman Brothers in the United States.

⁸If the funds are being used to finance long-term fixed-rate loans and interest rates rise, net interest margins are squeezed. But this risk can be hedged with instruments such as interest rate swaps (see Example 5.1).

Table 26.4 ASF Factors for Net Stable Funding Ratio

ASF Factor	Category
100%	Tier 1 and Tier 2 capital
90%	Preferred stock and borrowing with a remaining maturity greater than one year “Stable” demand deposits and term deposits with a remaining maturity less less than one year provided by retail or small business customers
80%	“Less Stable” demand deposits and term deposits with a remaining maturity less than one year provided by retail or small business customers
50%	Wholesale demand deposits and term deposits with a remaining maturity less than one year provided by non-financial corporates, sovereigns, central banks, multilateral development banks, and public-sector entities
0%	All other liability and equity categories

Basel III has introduced requirements involving two liquidity ratios that are designed to ensure that banks can survive liquidity pressures. The ratios are:

1. Liquidity Coverage Ratio (LCR); and
2. Net Stable Funding Ratio (NSFR)

The LCR focuses on a bank's ability to survive a 30-day period of liquidity disruptions. It is defined as:

$$\frac{\text{High-Quality Liquid Assets}}{\text{Net Cash Outflows in a 30-Day Period}}$$

The 30-day period considered in the calculation of this ratio is one of acute stress involving a downgrade of the bank's debt by three notches (e.g., from AA– to A–), a partial loss of deposits, a complete loss of wholesale funding, increased haircuts on secured funding (so that instruments posted as collateral are not valued as highly), and drawdowns on lines of credit. The Basel III regulations require the ratio to be greater than 100% so that the bank's liquid assets are sufficient to survive these pressures.

The NSFR focuses on liquidity management over a period of one year. It is defined as

$$\frac{\text{Amount of Stable Funding}}{\text{Required Amount of Stable Funding}}$$

The numerator is calculated by multiplying each category of funding (capital, wholesale deposits, retail deposits, etc.) by an available stable funding (ASF) factor, reflecting their stability. As shown in Table 26.4, the ASF for wholesale deposits is less than that for retail deposits, which in turn is smaller than that for Tier 1 or Tier 2 capital. The denominator is calculated from the items requiring funding. Each category of these is multiplied by a required stable funding (RSF) factor to reflect the permanence of the funding required. Some of the applicable factors are indicated in Table 26.5.

Table 26.5 RSF Factors for Net Stable Funding Ratio

RSF Factor	Category
0%	Cash
	Short-term instruments, securities, loans to financial entities if they have a residual maturity of less than one year
5%	Marketable securities with a residual maturity greater than one year if they are claims on sovereign governments or similar bodies with a 0% risk weight
20%	Corporate bonds with a rating of AA– or higher and a residual maturity greater than one year
	Claims on sovereign governments or similar bodies with a risk weight of 20%
50%	Gold, equity securities, bonds rated A+ to A–
65%	Residential mortgages
85%	Loans to retail and small business customers with a remaining maturity less than one year
100%	All other assets

Basel III requires the NSFR to be greater than 100% so that the calculated amount of stable funding is greater than the calculated required amount of stable funding.

Example 26.1

A bank has the following balance sheet:

Cash	5	Retail Deposits (stable)	40
Treasury Bonds (> 1 yr)	5	Wholesale Deposits	48
Mortgages	20	Tier 2 Capital	4
Small Business Loans	60	Tier 1 Capital	8
Fixed Assets	<u>10</u>		<u> </u>
	100		100

The Amount of Stable Funding is

$$40 \times 0.9 + 48 \times 0.5 + 4 \times 1.0 + 8 \times 1.0 = 72$$

The Required Amount of Stable Funding is

$$5 \times 0 + 5 \times 0.05 + 20 \times 0.65 + 60 \times 0.85 + 10 \times 1.0 = 74.25$$

The NSFR is therefore

$$\frac{72}{74.25} = 0.970$$

or 97.0%. The bank does not therefore satisfy the NSFR requirement.

The new rules are tough and have led to big changes in bank balance sheets. It was estimated in September 2014 that the rules will require U.S. banks to hold an extra \$100 billion of liquid assets.

26.2.6 Counterparty Credit Risk

For each of its derivatives counterparties, a bank calculates a *credit value adjustment* (CVA) which has been discussed in earlier chapters. This is the expected loss because of the possibility of a default by the counterparty. The way in which the calculation of CVA is carried out was described in Chapter 18. Reported profit is reduced by the total of the CVAs for all counterparties.

The CVA for a counterparty can change because either (a) the risk factors underlying the value of the derivatives entered into with the counterparty change or (b) the credit spreads applicable to the counterparty's borrowing change. Basel III requires the CVA risk arising from changing credit spreads to be a component of market risk capital. Once CVA has been calculated, it is a relatively simple matter to calculate the delta and gamma with respect to a parallel shift in the term structure of the counterparty's credit spread. These can be used to calculate appropriate market risk measures.

The Basel Committee proposed changes to the calculation of capital for CVA risk in 2015.⁹ A major objective was to allow banks that hedged the impact of both credit spread changes and risk factor changes on CVA to take both into account when calculating regulatory capital. Without this change, hedges for risk factor changes would increase capital charges.

26.2.7 G-SIBs, SIFIs, and D-SIBs

Regulators are particularly concerned that large, systemically important financial institutions keep sufficient capital to avoid a repeat of the government bailouts during the 2007 to 2009 financial crisis.

The term G-SIB stands for global systemically important bank, whereas the term SIFI (systemically important financial institution) is used to describe both banks and nonbanks that are considered to be systemically important. The popular view of SIFIs is that they are “too big to fail,” and have been identified as the financial institutions that will have to be bailed out if they run into financial difficulties.

The systemic importance of a bank or other financial institution depends on the effect that its failure could have on the global financial system. This in turn depends on the nature of its activities and the contracts it has entered into with other financial institutions globally. The Basel Committee uses a scoring methodology to determine which banks are G-SIBs. Other approaches involving network theory have been attempted by some researchers.

In 2013, the Basel Committee published the final version of rules that call for more Tier 1 equity capital for G-SIBs.¹⁰ G-SIBs are categorized according to whether the extra equity capital is 1%, 1.5%, 2%, 2.5%, or 3.5% of risk-weighted assets.

⁹See Basel Committee on Banking Supervision, “Review of Credit Valuation Adjustment Risk Framework,” October 1, 2015.

¹⁰See Basel Committee on Banking Supervision, “Global Systemically Important Banks: Updated Assessment Methodology and the Higher Loss Absorbency Requirement,” July 2013.

A list of G-SIBs is published by the Financial Stability Board annually. In November 2021, the list included the 30 banks shown in Table 26.6. A total of 18 banks were in the 1% category, eight were in the 1.5% category, three were in the 2% category, and one (JP Morgan Chase) was in the 2.5% category. None were in the 3.5% category. The CET1 capital, as a percent of risk-weighted assets required by the banks in Table 26.6, was therefore the baseline 4.5% plus 2.5% for the capital conservation buffer plus the amount indicated in the table plus any extra amounts required by national supervisors. The CET1 plus AT1 required capital was 1.5% more than this.

In November 2015, the Financial Stability Board issued proposals concerning the total loss-absorbing capacity (TLAC) of G-SIBs. The proposals were developed in consultation with the Basel Committee on Banking Supervision and were in response to a call by G20 leaders at the 2013 St. Petersburg Summit. TLAC consists of instruments that can absorb losses and protect depositors in the event of a default. The instruments include equity, preferred shares, and subordinated debt. From January 1, 2019, the minimum TLAC for G-SIBs was set at 16% of risk-weighted assets, increasing to 18% on January 1, 2022. It also had to be less than 6% of the leverage ratio denominator on January 1, 2019, increasing to 6.75% on January 1, 2022. Non-compliance with TLAC could impede a bank's ability to issue dividends.

Some countries are requiring greater capital than the minimums set by the Basel Committee and the Financial Stability Board or are implementing the rules earlier than required. An example of a country requiring that its banks hold high amounts of capital is Switzerland, which has two G-SIBs (UBS and Credit Suisse). The high capital requirements are understandable because the banks are large in relation to the Swiss economy, so that a failure of either of them would be catastrophic.

National regulators designate some banks that have not been classified as G-SIBs as domestic systemically important banks (D-SIBs). These banks may be subject to capital requirements higher than the minimum, extra disclosure requirements, or stringent stress tests.

Table 26.6 G-SIB Banks in 2021, Categorized According to Extra Capital Required

Extra Capital	Banks
1%	Agricultural Bank of China, Bank of New York Mellon, Credit Suisse, Groupe BPCE, Groupe Crédit Agricole, ING Bank, Mizuho FG, Morgan Stanley, Royal Bank of Canada, Santander, Société Générale, Standard Chartered, State Street, Sumitomo Mitsui FG, Toronto Dominion, UBS, UniCredit, Wells Fargo
1.5%	Bank of America, Bank of China, Barclays, China Construction Bank, Deutsche Bank, Goldman Sachs, Industrial and Commercial Bank of China, Mitsubishi UFJ FG
2%	BNP Paribas, Citigroup, HSBC
2.5%	JP Morgan Chase

26.3 Contingent Convertible Bonds

An interesting development in the capitalization of banks has been what are known as *contingent convertible bonds* (CoCos). Traditionally, convertible bonds have been bonds issued by a company where, in certain circumstances, the holder can choose to convert them into equity at a predetermined exchange ratio. Typically the bond holder chooses to convert when the company is doing well and the stock price is high. CoCos are different in that they automatically get converted into equity when certain conditions are satisfied. Typically, these conditions are satisfied when the company is experiencing financial difficulties. CoCos provide a higher rate of interest than regular debt.

CoCos are attractive to banks because in normal times the bonds are debt and allow the bank to report a relatively high return on equity. When the bank experiences financial difficulties and incurs losses, the bonds are converted into equity and the bank is able to continue to maintain an equity cushion and avoid insolvency. From the point of view of regulators, CoCos are potentially attractive because they avoid the need for a bailout. Indeed, the conversion of CoCos is sometimes referred to as a “bail-in.” New equity for the financial institution is provided from within by private-sector bondholders rather than from outside by the public sector.

A key issue in the design of CoCos is the specification of the trigger that forces conversion and the way that the exchange ratio (number of shares received in exchange for one bond) is set. A popular trigger in the bonds issued so far is the ratio of Tier 1 equity capital to risk-weighted assets. Another possible trigger is the ratio of the market value of equity to book value of assets.

Lloyd's Banking Group, Rabobank Netherlands, and Credit Suisse were among the first banks to issue CoCos. Business Snapshot 26.1 provides a description of the bonds issued by Credit Suisse in 2011. These bonds get converted into equity if either Tier 1 equity capital falls below 7% of risk-weighted assets or the Swiss bank regulators determine that the bank requires public-sector support. CoCos have been popular in Europe with over \$200 billion of issues by 2020.

CoCos (prior to conversion) qualify as additional Tier 1 capital if the trigger, defined in terms of the ratio of Tier 1 equity capital to risk-weighted assets, is set at 5.125% or higher. Otherwise they qualify as Tier 2 capital.

26.4 Use of Standardized Approaches and SA-CCR

In December 2017, the Basel Committee announced that, starting in 2022, it would require the implementation of a standardized approach for all capital calculations. A bank's total capital requirement will be the maximum of (a) that calculated as before using approved internal models and (b) a certain percentage of that given by the standardized approaches. The percentage will 50% in 2023, rising to 72.5% in 2028.

BUSINESS SNAPSHOT 26.1

Credit Suisse's CoCo Bond Issues

On February 14, 2011, Credit Suisse announced that it had agreed to exchange \$6.2 billion of existing investments by two Middle Eastern investors, Qatar Holding LLC and the Olayan Group LLC, for CoCo bonds. The bonds automatically convert into equity if either of the following two conditions are satisfied:

1. The Tier 1 equity capital of Credit Suisse falls below 7% of risk-weighted assets.
2. The Swiss bank regulator determines that Credit Suisse requires public-sector support to prevent it from becoming insolvent.

Credit Suisse followed this announcement on February 17, 2011, with a public issue of \$2 billion of CoCos. These securities have similar terms to ones held to the Middle Eastern investors and were rated BBB+ by Fitch. They mature in 2041 and can be called any time after August 2015. The coupon is 7.875%. Any concerns that the market had no appetite for CoCos were alleviated by this issue. It was 11 times oversubscribed.

Credit Suisse indicated that it plans to satisfy one-third of the non-equity capital requirement with bonds similar to those just described and two-thirds of the non-equity capital requirement with bonds where the conversion trigger is about 5% (rather than 7%) of risk-weighted assets.

The standardized approach for market risk capital applicable in 2022 will be explained in Chapter 27. As explained in Chapter 20, operational risk capital will be based entirely on a standardized approach (the SMA). The standardized risk weights for credit risk are a revision of the Basel II weights.

The standardized approach for calculating exposure at default (EAD) for derivatives in Basel I has been improved with what is known as SA-CCR. EAD is

$$1.4 \times (RC + PFE)$$

where RC is the replacement cost and PFE is potential future exposure. For transactions where no collateral is posted, the RC is calculated as

$$RC = \max(V, 0)$$

For transactions that are subject to collateral agreements this is adjusted to

$$RC = \max(V - C, D, 0)$$

where C is the net value of the collateral currently posted by the counterparty (with the collateral posted by the bank being considered negative). The variable D is the change in the V that could occur without any additional collateral being received (e.g., because of minimum transfer amounts in the collateral agreement). For example, if $V = C = 50$ and V could increase to 51 without any additional collateral being received, $D = 1$ and the $RC = 1$. In the case of non-cash collateral, a haircut is applied in the calculation of C .

The PFE is a relatively sophisticated way of calculating the add-on to the current exposure. Two factors taken into account are:

1. When the counterparty has posted excess collateral, risk is reduced and the reduction in risk increases as the amount of the excess collateral increases.
2. When no collateral is posted, credit exposure reduces as V becomes more negative. The PFE add-on amount when V is slightly negative should be greater than when V is highly negative.

26.5 Dodd–Frank Act

The Dodd–Frank Act in the United States was signed into law in July 2010. Its aim was to prevent future bailouts of financial institutions and protect the consumer. A summary of the main regulations is as follows:

1. Two new bodies, the Financial Stability Oversight Council (FSOC) and the Office of Financial Research (OFR), were created to monitor systemic risk and research the state of the economy. Their tasks are to identify risks to the financial stability of the United States, promote market discipline, and maintain investor confidence.
2. The orderly liquidation powers of the Federal Deposit Insurance Corporation (FDIC) were expanded. The Office of Thrift Supervision was eliminated.
3. The amount of deposits insured by the FDIC was increased permanently to \$250,000. (Previously, the \$250,000 limit had been regarded as temporary.)
4. Regulations were introduced requiring large hedge funds and similar financial intermediaries to register with the SEC and report on their activities.
5. A Federal Insurance Office was created to monitor all aspects of the insurance industry and work with state regulators.
6. Proprietary trading and other similar activities of deposit-taking institutions were curtailed. This is known as the “Volcker rule” because it was proposed by former Federal Reserve chairman Paul Volcker.
7. Some high-risk trading operations were required to be spun off into separately capitalized affiliates.
8. Standardized over-the-counter derivatives had to be traded on electronic platforms known as swap execution facilities (SEFs). These are similar to exchanges and are designed to make prices in the OTC market more transparent. Standardized over-the-counter derivatives between financial institutions had to be cleared by

central clearing parties (CCPs). (See Chapter 6 for a further discussion of CCPs.) The Commodity Futures Trading Commission (CFTC) was given responsibility for monitoring the activities of CCPs and SEFs.

9. The Federal Reserve was required to set risk management standards for systemically important financial utilities engaged in activities such as payment, settlement, and clearing.
10. Protection for investors was increased and improvements were made to the regulation of securities.
11. Rating agencies were required to make the assumptions and methodologies behind their ratings more transparent and the potential legal liabilities of rating agencies were increased. An Office of Credit Ratings was created at the SEC to provide oversight of rating agencies.
12. The use of external credit ratings in the regulation of financial institutions was discontinued. (This provision of the Act brings Dodd–Frank into direct conflict with the Basel Committee, which, as we have seen in this chapter and the last one, does make some use of external credit ratings.)
13. A Consumer Financial Protection Bureau was created within the Federal Reserve to ensure that consumers get clear and accurate information when they shop for financial products, such as mortgages and credit cards.
14. Issuers of securitized products were required (with some exceptions) to own 5% of each product created.
15. Federal bank regulators were required to issue regulations that discourage the use of compensation arrangements that might lead to excessive risk taking (e.g., compensation arrangements based on short-run performance). Shareholders were given a non-binding vote on executive compensation. A requirement that board compensation committees be made up of independent directors was introduced.
16. Mortgage lenders were required to make a reasonable good faith determination based on verified and documented information that the borrower has the ability to repay a loan. Failure to do this might lead to a foreclosure being disallowed.
17. Large financial firms were required to have board committees where at least one expert has risk management experience at a large complex firm.
18. The FDIC was allowed to take over a large financial institution when it failed and sell its assets, imposing losses on shareholders and creditors, with the costs of failures being paid for by the financial industry.
19. FSOC and OFR, which as mentioned earlier had the responsibility of monitoring systemic risk, were charged with identifying systemically important financial institutions (SIFIs).
20. The Federal Reserve Board and the FDIC required all SIFIs to prepare what is known as *living wills*, mapping out how they could raise funds in a crisis and how their operations would be wound up in the event of failure.

The Dodd–Frank Act did not define a future role for Fannie Mae and Freddie Mac, which were key players in the U.S. mortgage market. These agencies were taken over by

the U.S. government in September 2008. The Dodd–Frank Act has been loosened somewhat since 2010. For example, in 2018, small- and medium-sized banks were exempted from some regulations. In 2020, banks were granted permission to invest in venture capital funds.

26.6 Legislation in Other Countries

The large banks are truly global and when regulations vary throughout the world, they are liable to move all or part of their operations from one jurisdiction to another to obtain more favorable treatment. Although all countries are subject to the same Basel III rules, local regulators have some discretion in the application of rules, and legislation is not the same across all countries. In 2011, the Swiss bank UBS made headlines by suggesting that it might move its investment bank headquarters from Zurich to London, Singapore, or New York to avoid the higher capital requirements imposed by Swiss regulators. (It did not do this.)

The previous section outlined the rules introduced by legislators in the United States. Legislation in other countries has addressed some of the same issues. In the United Kingdom, an independent committee was chaired by Sir John Vickers to consider issues facing the banking industry, and new legislation in the form of the Financial Services (Banking Reform) Act 2013 was passed. In the European Union a committee headed by Erkki Liikanen was set up in November 2011 and published its report in October 2012.

Some of the rules and recommendations in the United Kingdom, European Union, and other countries are similar to those in the United States. For example, the Dodd–Frank Act requires originators of securitized products in the United States to keep 5% of all assets created. (See item 14 of the previous list.) A similar provision exists in the Capital Requirement Directive 2 (CRD2) of the European Union.¹¹ Most national legislators have agreed that standardized over-the-counter derivatives should be cleared through central clearing houses instead of being cleared bilaterally (see list item 8).

Sometimes the rules in different countries are similar but not exactly the same. The Dodd–Frank Act included the requirement that standardized derivatives be traded on swap execution facilities (see list item 8). The European Union has introduced similar trading platforms, known as organized trading facilities (OTFs).

The most controversial aspect of Dodd–Frank is probably the Volcker rule (see item 6), which prohibits banks from proprietary trading and restricts investment in hedge funds and private equity companies by banks and their affiliates. The rationale for the rule is reasonable: Banks should not be allowed to speculate with depositors' funds, because those funds are insured by the FDIC. The rule may be difficult to enforce, though, because for a bank with a large portfolio of positions, it can be difficult to tell whether a particular new trade is entered into for hedging or speculative purposes.¹² In spite of this, many U.S. banks have closed down their proprietary trading desks, and some

¹¹However, Germany has increased the 5% to 10%.

¹²A joke circulating on Wall Street was that banks would have to employ psychologists or psychiatrists to determine what was going on in the mind of the trader at the time of a trade!

employees working on those desks have left to join hedge funds. The Vickers committee in the United Kingdom recognized the importance of protecting depositors' funds but did not go so far as to prohibit proprietary trading. The retail operations of banks had (by 2019) to be ring-fenced to insulate them from riskier activities such as trading and investment banking. The Liikanen committee in the European Union similarly requires a separation of core banking activities from proprietary trading and other high-risk activities.

Regulators in most countries consider living wills (see list item 20) to be important for SIFIs and are applying pressure on SIFIs to develop them. SIFIs tend to have developed complex organizational structures for tax and regulatory purposes. The living will requirement may result in this being simplified so that the different activities of a SIFI are in separately capitalized legal entities, not all of which need to be bailed out in the event of the SIFI's failure. Regulators have the option of forcing SIFIs to divest certain operations, or even break up entirely, if their living wills are deemed unsatisfactory. They hope that living wills avoid a replay of the panic and cross-border litigation that erupted when the Lehman Brothers investment bank collapsed in 2008. They also think that the multiyear litigation concerned with the termination of Lehman's derivatives portfolio could have been avoided if Lehman and its ISDA master agreements had been structured so that it was possible to keep the portfolio alive and to unwind it in an orderly way in the few days after bankruptcy.

Compensation is an important issue. Pre-crisis, the annual bonus was a large part of the compensation for many traders and other employees and led them to have a relatively short-term horizon in their decision making. If losses were incurred after the payment of a bonus, they did not have to return the bonus. Many banks have recognized the problem and voluntarily moved to bonuses that are deferred by being spread out over three to five years, rather than all being paid in one year. If a trader shows good results in one year and bad results in the next, some of the bonus applicable to the good year will be deferred and then "clawed back" during the bad year. The Dodd-Frank Act restrictions on pay in the financial sector are relatively mild. When financial institutions received funds during the crisis under the Troubled Asset Relief Program (TARP), compensation was restricted. But, as soon as the funds were paid back, banks had much more freedom in their compensation arrangements.¹³

Some other countries have restricted compensation. Sometimes the restriction is temporary. For example, in 2009 the United Kingdom introduced a one-time "supertax" on bonuses in excess of £25,000. But in other cases it is permanent. For example, in April 2019, the European Union introduced CRD 5, which confirmed a cap on the bonuses on bankers classified as "risk takers." Under this directive the maximum ratio of bonus to fixed pay is 1.0 (with some flexibility to increase that ratio to 2.0 with shareholder approval).

¹³Indeed, this was a major incentive to repay the TARP money as quickly as possible!

Summary

The financial crisis that started in 2007 was the worst that many parts of the world had seen since the 1930s. Some financial institutions failed. Others had to be bailed out with taxpayers' money. Not surprisingly, both the Basel Committee and national governments decided that a major overhaul of the regulations affecting financial institutions was required.

The Basel II.5 regulations increased the capital banks were required to keep for market risk. They recognized that capital should reflect the volatilities and correlations experienced during stressed market conditions as well as during normal market conditions; they eliminated some of the ways banks could reduce regulatory capital by moving items from the banking book to the trading book; and they created a special capital requirement for derivatives dependent on credit correlation, which had been a particular problem during the crisis.

Basel III dramatically increased the amount of equity capital banks were required to keep. It also recognized that many of the problems of banks during the crisis were liquidity problems and imposed new liquidity requirements for financial institutions.

National governments have also introduced new rules for financial institutions. In the United States, the Dodd–Frank Act has many provisions designed to protect consumers and investors, avoid future bailouts, and monitor the functioning of the financial system more carefully. Similar, though not necessarily identical, regulations exist in other countries.

One problem facing regulators is what are referred to as *unintended consequences*. Basel I had the unintended consequence of discouraging loans to high-quality corporations because of the 100% risk weight that would be assigned. The 1996 Amendment and the development of the credit derivatives market that came after it encouraged banks to find ways of moving credit risks from the banking book to the trading book in order to reduce capital requirements. There will no doubt be unintended consequences of Basel III and the legislation. Hopefully, the benefits of the new measures will outweigh any harm to the financial system arising from the unintended consequences.

Further Reading

Acharya, V. V., T. F. Cooley, M. P. Richardson, and I. Walter. *Regulating Wall Street: The Dodd–Frank Act and the New Architecture of Global Finance*. Hoboken, NJ: John Wiley & Sons, 2011.

Bank for International Settlements. “Basel III: Finalizing Post-Crisis Reforms,” December 2017.

Basel Committee on Banking Supervision. “Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems,” June 2011.

Basel Committee on Banking Supervision. “Basel III: International Framework for Liquidity Risk Measurement Standards and Monitoring,” December 2010.

Basel Committee on Banking Supervision. “Guidelines for Computing Capital for Incremental Risk in the Trading Book,” July 2009.

Basel Committee on Banking Supervision. “Revisions to the Basel II Market Risk Framework,” February 2011.

Finger, C. “CreditMetrics and Constant Level of Risk.” MSCI, 2010.

Practice Questions and Problems (Answers at End of Book)

- 26.1 What are the three major components of Basel II.5?
- 26.2 What are the six major components of Basel III?
- 26.3 What is the difference between VaR as it has been traditionally measured and stressed VaR?
- 26.4 Explain how the incremental risk charge is calculated. Why was it introduced by the Basel Committee?
- 26.5 What is the difference between the capital required for a AAA-rated ABS with principal of \$100 million and a AAA-rated ABS CDO with a principal of \$100 million using the standardized approach?
- 26.6 By how much has the Tier 1 equity capital (including the capital conservation buffer) increased under Basel III compared with the Tier 1 equity capital requirement under Basel I and II?
- 26.7 Suppose that the Tier 1 equity ratio for a bank is 6%. What is the maximum dividend, as a percent of earnings, that can be paid if (a) there is no countercyclical buffer and (b) there is a 2.5% countercyclical buffer?
- 26.8 Explain how the leverage ratio differs from the usual capital ratios calculated by regulators.
- 26.9 Explain how the liquidity coverage ratio and the net stable funding ratio are defined.
- 26.10 How would the net stable funding ratio in Example 26.1 change if half the wholesale deposits were replaced by stable retail deposits?
- 26.11 What is CVA? What new regulations concerning CVA were introduced in Basel III?
- 26.12 Explain how CoCo bonds work. Why are they attractive to (a) banks and (b) regulators?

Further Questions

- 26.13 Explain one way that the Dodd–Frank Act is in conflict with (a) the Basel international regulations and (b) the regulations introduced by other national governments.

26.14 A bank has the following balance sheet:

Cash	3	Retail Deposits (stable)	25
Treasury Bonds (>1 yr)	5	Retail Deposits (less stable)	15
Corporate Bonds Rated A	4	Wholesale Deposits	44
Mortgages	18	Preferred Stock (>1 yr)	4
Small Business Loans (<1 yr)	60	Tier 2 Capital	3
Fixed Assets	<u>10</u>	Tier 1 Capital	<u>9</u>
	100		100

- (a) What is the Net Stable Funding Ratio?
- (b) The bank decides to satisfy Basel III by raising more (stable) retail deposits and keeping the proceeds in Treasury bonds. What extra retail deposits need to be raised?

Chapter 27

Fundamental Review of the Trading Book

In May 2012, the Basel Committee on Banking Supervision issued a consultative document proposing major revisions to the way regulatory capital for market risk is calculated. This is referred to as the “Fundamental Review of the Trading Book” (FRTB).¹ The Basel Committee then followed its usual process of requesting comments from banks, revising the proposals, and carrying out Quantitative Impact Studies (QISs).² The final version of the rules was published by the Basel Committee in January 2019.³ The internationally agreed implementation date is January 2023, but many jurisdictions indicated that they will be a year or two later than that.

FRTB’s approach to determining capital for market risk is much more complex than the approaches previously used by regulators. The purpose of this chapter is to outline its main features.

¹See Bank for International Settlements, “Consultative Document: Fundamental Review of the Trading Book,” May 2012.

²QISs are calculations carried out by banks to estimate the impact of proposed regulatory changes on capital requirements.

³See Bank for International Settlements, “Minimum Capital Requirements for Market Risk,” January 2019.

27.1 Background

The Basel I calculations of market risk capital were based on a value at risk (VaR) calculated for a 10-day horizon with a 99% confidence level. The VaR was “current” in the sense that calculations made on a particular day were based on the behavior of market variables during an immediately preceding period of time (typically, one to four years). Basel II.5 required banks to calculate a “stressed VaR” measure in addition to the current measure. As explained in Sections 12.1 and 26.1, this is VaR where calculations are based on the behavior of market variables during a 250-day period of stressed market conditions. To determine the stressed period, banks were required to go back through time searching for a 250-day period where the observed movements in market variables would lead to significant financial stress for the current portfolio.

FRTB changes the measure used for determining market risk capital. Instead of VaR with a 99% confidence level, it uses expected shortfall (ES) with a 97.5% confidence level. The measure is actually stressed ES with a 97.5% confidence. This means that, as in the case of stressed VaR, calculations are based on the way market variables have been observed to move during stressed market conditions.

For normal distributions, VaR with a 99% confidence and ES with a 97.5% confidence are almost exactly the same. Suppose losses have a normal distribution with a mean μ and standard deviation σ . The 99% VaR is $\mu + 2.326\sigma$ while the 97.5% expected shortfall is $\mu + 2.338\sigma$.⁴ (See Problem 27.2.) For non-normal distributions, they are not equivalent. When the loss distribution has a heavier tail than a normal distribution, the 97.5% ES can be considerably greater than the 99% VaR.

Under FRTB, the 10-day time horizon used in Basel I and Basel II.5 is changed to reflect the liquidity of the market variable being considered. FRTB considers changes to market variables that would take place (in stressed market conditions) over periods of time reflecting their liquidity. The changes are referred to as *shocks*. The market variables are referred to as *risk factors*. The periods of time considered are referred to as *liquidity horizons*. Five different liquidity horizons are specified: 10 days, 20 days, 40 days, 60 days, and 120 days. The allocation of risk factors to these liquidity horizons is indicated in Table 27.1.

FRTB specifies both a standardized approach and an internal models approach for calculating market risk capital. Even when banks have been approved to use the internal models approach, they are required by regulators to calculate required capital under both approaches. This is consistent with the Basel Committee’s plans to use standardized approaches to provide a floor for capital requirements. As discussed in Section 26.4, in December 2017, the Basel Committee announced a move to a situation where total required capital is at least 72.5% of that given by standardized approaches. It will achieve

⁴From equation (11.2), the ES for a normal distribution with mean μ and standard deviation σ is $\mu + \sigma \exp(-Y^2/2)/[\sqrt{2\pi}(1-X)]$ where X is the confidence level and Y is the point on a normal distribution that has a probability of $1-X$ of being exceeded. This can also be written $\mu + \sigma^2 f(VaR)/(1-X)$ where f is the probability density function for the loss.

Table 27.1 Allocation of Risk Factors to Liquidity Horizons

Risk Factor	Horizon (days)
Interest rate (dependent on currency)	10–60
Interest rate volatility	60
Credit spread: sovereign, investment grade	20
Credit spread: sovereign, non-investment grade	40
Credit spread: corporate, investment grade	40
Credit spread: corporate, non-investment grade	60
Credit spread: other	120
Credit spread volatility	120
Equity price: large cap	10
Equity price: small cap	20
Equity price: large cap volatility	20
Equity price: small cap volatility	60
Equity: other	60
Foreign exchange rate (dependent on currency)	10–40
Foreign exchange volatility	40
Energy price	20
Precious metal price	20
Other commodities price	60
Energy price volatility	60
Precious metal volatility	60
Other commodities price volatility	120
Commodity (other)	120

this by 2028 with a five-year phase-in period. These changes are a culmination of a trend by the Basel Committee since the 2008 crisis to place less reliance on internal models and to use standardized models to provide a floor for capital requirements.

A difference between FRTB and previous market risk regulatory requirements is that most calculations are carried out at the trading desk level. Furthermore, permission to use the internal models approach is granted on a desk-by-desk basis. Therefore it is possible that, at a particular point in time, a bank's foreign currency trading desk has permission to use the internal models approach while the equity trading desk does not.

In earlier chapters, we saw how the ways in which capital is calculated for the trading book and the banking book are quite different. This potentially gives rise to regulatory arbitrage where banks choose to allocate instruments to either the trading book or the banking book so as to minimize capital. In Basel II.5, the incremental risk charge made this less attractive. FRTB counteracts regulatory arbitrage by defining more clearly than previously the differences between the two books.

27.2 Standardized Approach

Under the standardized approach, the capital requirement is the sum of three components: a risk charge calculated using a risk sensitivity approach, a default risk charge, and a residual risk add-on.

Consider the first component. Seven risk classes (corresponding to trading desks) are defined (general interest rate risk, foreign exchange risk, commodity risk, equity risk, and three categories of credit spread risk). Within each risk class, a delta risk charge, vega risk charge, and curvature risk charge are calculated.

The delta risk charge for a risk class is calculated using the risk weights and weighted sensitivity approach described in Section 13.6:

$$\text{Delta Risk Charge} = \sum_i \sum_j \rho_{ij} \delta_i \delta_j W_i W_j \quad (27.1)$$

where the summations are taken over all risk factors in the risk class. The risk weights, W_i , and the correlations between risk factors, ρ_{ij} , are determined by the Basel Committee.⁵ The sensitivities (or deltas), δ_i , are determined by the bank. As explained in Chapter 13, in the case of risk factors such as equity prices, exchange rates, or commodity prices, the deltas measure the sensitivity of the portfolio to percentage changes. For example, if a 1% increase in a commodity price would increase the value of a portfolio by \$3,000, the delta would be $3,000/0.01 = 300,000$. In the case of risk factors such as interest rates and credit spreads, the deltas are defined in terms of absolute changes. For example, if the effect of an interest rate increasing by one basis point (0.0001) is to reduce the value of a portfolio by \$200, the delta with respect to that interest rate would be $-200/0.0001 = -2,000,000$.

Consider how the risk weights, W_i , might be set by regulators. Suppose first that all risk factors are equity prices, exchange rates, or commodity prices, so the deltas are sensitivities to percentage changes. If W_i were set equal to the daily volatility of risk factor i for all i , the risk charge in equation (27.1) would equal the standard deviation of change in the value of the portfolio per day. This corresponds to the Markowitz result discussed in Chapter 1. If W_i were set equal to the daily volatility of risk factor i in stressed market conditions (the stressed daily volatility) for all i , equation (27.1) would give the standard deviation of the daily change of the portfolio in stressed market conditions. In practice, the W_i are set equal to multiples of the stressed daily volatility to reflect the liquidity horizon and the confidence level that regulators wish to consider. Suppose that the stressed daily volatility of risk factor i is estimated as 2% and that the risk factor has a 20-day liquidity horizon. The risk weight might be set as $0.02 \times \sqrt{20} \times 2.338 = 0.209$. (Note that the 2.338 multiplier reflects the amount by which a standard deviation has to be multiplied to get ES with a 97.5% confidence when a normal distribution is assumed.)

Now suppose that the risk factors are interest rates and credit spreads so that deltas are sensitivities with respect to actual changes measured in basis points. The W_i for risk factor

⁵Banks are required to test the effect of multiplying the correlations specified by the Basel Committee by 1.25, 1.00, and 0.75 and then set the capital charge equal to the greatest result obtained.

i is set equal to a multiple of the stressed daily standard deviation for all i . If the multiple were 1, the formula would give the standard deviation of the value of the portfolio in one day. In practice the multiple is determined as just described to reflect the liquidity horizon and confidence level.

Vega risk is handled similarly to delta risk.⁶ A vega risk charge is calculated for each risk class using equation (27.1). The risk factors (counted by the i and j) are now volatilities. The summation is taken over all volatilities in the risk class. The parameter δ_i is actually a vega. It is the sensitivity of the value of the portfolio to small changes in volatility i .⁷ The parameter ρ_{ij} is the correlation between changes in volatility i and volatility j , and W_i is the risk weight for volatility i . The latter is determined similarly to the delta risk weights to reflect the volatility of the volatility i , its liquidity horizon, and the confidence level.

There are assumed to be no diversification benefits between risk factors in different risk classes and between the vega risks and delta risks within a risk class. The end product of the calculations we have described so far is therefore the sum of the delta risk charges across the seven risk classes plus the sum of the vega risk charges across the seven risk classes.

27.2.1 Term Structures

In the case of risk factors such as interest rates, volatilities, and credit spreads, there is usually a term structure defined by a number of points. For example, an interest rate term structure is sometimes defined by 10 points. These are the zero-coupon interest rates for maturities of 3 months, 6 months, 1 year, 2 years, 3 years, 5 years, 10 years, 15 years, 20 years, and 30 years. As discussed in Chapter 13, each vertex of the term structure is a separate risk factor for the purposes of using equation (27.1). The delta of a portfolio with respect to a one basis point move in one of the vertices on the term structure is calculated by increasing the position of the vertex by one basis point while making no change to the other vertices. The Basel Committee defines risk weights for each vertex of the term structure and correlations between the vertices of the same term structure.

A simplification is used when correlations between points on different term structures are defined. The correlations between point A on term structure 1 and point B on term structure 2 are assumed to be the same for all A and B. This was one of the alternatives considered in Chapter 14; see equation (14.2).

27.2.2 Curvature Risk Charge

The curvature risk charge is a capital charge for a bank's gamma risk exposure under the standardized approach. Consider the exposure of a portfolio to the i th risk factor. Banks are required to test the effect of increasing and decreasing the risk factor by its

⁶As discussed in Chapter 13, this works well because most of the value of a derivative is in many cases approximately linearly dependent on volatility.

⁷Banks can choose whether it is percentage or actual changes in volatility that are considered.

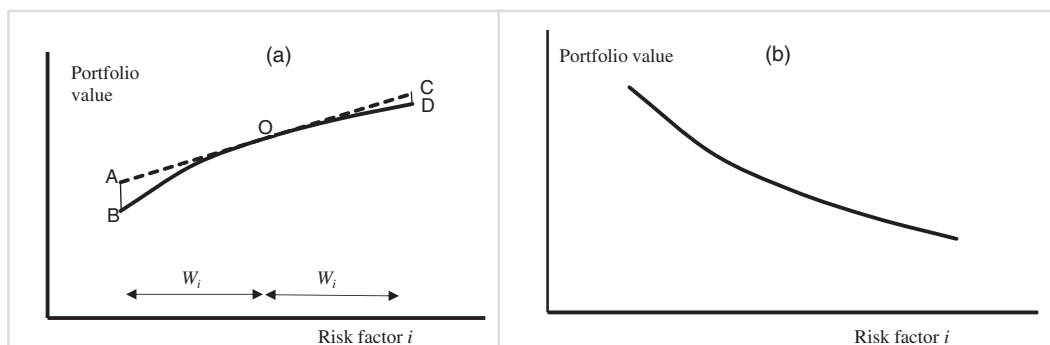


Figure 27.1 Calculation of Curvature Risk Charge for a Risk Factor
In Figure 27.1a, the curvature risk charge is AB ; in Figure 27.1b, it is zero.

risk weight, W_i . If the portfolio is linearly dependent on the risk factor, the impact of an increase of W_i in the risk factor is $W_i\delta_i$. Similarly, the impact of a decrease of W_i in the risk factor is $-\delta_i W_i$. To evaluate the impact of curvature net of the delta effect, the standardized approach therefore calculates

1. $W_i\delta_i$ minus the impact of an increase of W_i in the risk factor, and
2. $-W_i\delta_i$ minus the impact of a decrease in the risk factor of W_i .

The curvature risk charge for the risk factor is the greater of these two. If the impact of curvature net of delta is negative, it is counted as zero. The calculation is illustrated in Figure 27.1. In Figure 27.1a, the portfolio value is currently given by point O . If there were no curvature, an increase of W_i in the risk factor would lead to the portfolio value at point C , whereas a decrease of W_i in the risk factor would lead to the portfolio value at point A . Because of curvature, an increase of W_i leads to the portfolio value at point D , and a decrease of W_i leads to the portfolio value at point B . Since $AB > CD$, the risk charge is AB . In Figure 27.1b, the risk charge is zero because curvature actually increases the value of the position (relative to what delta would suggest) for both increases and decreases in the risk factor. (Figure 27.1a could correspond to a short position in an option; Figure 27.1b could correspond to a long position in an option.)

When there are several risk factors, each is handled similarly to Figure 27.1. When there is a term structure (e.g., for interest rates, credit spreads, and volatilities), all points are shifted by the same amount for the purpose of calculating the effect of curvature. The shift is the largest W_i for the points on the term structure. In the case of an interest rate term structure, the W_i corresponding to the three-month vertex might be the largest W_i , so this would define an upward and downward parallel shift in the term structure. The delta effect is removed for each point on the term structure by using the δ_i for that point.

The curvature risk charges for different risk factors are combined to determine a total curvature risk charge. When diversification benefits are allowed, aggregation formulas broadly similar to those used for deltas are used with correlations specified by the Basel Committee.

27.2.3 *Default Risk Charge*

Risks associated with counterparty credit spread changes are handled separately from risks associated with counterparty defaults in FRTB. In the standardized approach, credit spread risks are handled using the delta/vega/curvature approach described earlier. Default risks, sometimes referred to as *jump-to-default* (JTD) risks, are handled by a separate default risk charge. This is calculated by multiplying each exposure by a loss given default (LGD) and a default risk weight. Both the LGD and the risk weight are specified by the Basel Committee. For example, the LGD for senior debt is specified as 75% and the default risk weight for a counterparty rated A is 3%. Equity positions are subject to a default risk charge with an LGD = 100%. Rules for offsetting exposures are specified.

27.2.4 *Residual Risk Add-On*

The residual risk add-on considers risks that cannot be handled by the delta/vega/curvature approach described earlier. It includes exotic options when they cannot be considered as linear combinations of plain vanilla options. The add-on is calculated by multiplying the notional amount of the transaction by a risk weight that is specified by the Basel Committee. In the case of exotic options the risk weight is 1%.

27.2.5 *A Simplified Approach*

In this section, we have described the standardized approach that the Basel Committee requires all large banks to use. In June 2017 the Basel Committee published a consultative document outlining a simplified standardized approach that it proposes for smaller banks.⁸ This has been included in the final January 2019 document. The full approach is simplified in a number of ways. For example, vega and gamma risk do not have to be considered. This should make FRTB more attractive to jurisdictions such as the United States that have many small banks that tend to enter into only relatively simple transactions.

27.3 **Internal Models Approach**

The internal models approach requires banks to estimate stressed ES with a 97.5% confidence. FRTB does not prescribe a particular method for doing this. Typically the historical simulation approach described in Chapter 12 is likely to be used. Risk factors are allocated to liquidity horizons as indicated in Table 27.1. Define:

Category 1 Risk Factors: Risk factors with a time horizon of 10 days

Category 2 Risk Factors: Risk factors with a time horizon of 20 days

Category 3 Risk Factors: Risk factors with a time horizon of 40 days

Category 4 Risk Factors: Risk factors with a time horizon of 60 days

Category 5 Risk Factors: Risk factors with a time horizon of 120 days

⁸See Basel Committee on Banking Supervision, “Simplified Alternative to the Standardized Approach to Market Risk Capital Requirements,” June 2017.

As we shall see, all calculations are based on considering 10-day changes in the risk factors. In Basel I and Basel II.5, banks are allowed to deduce the impact of 10-day changes from the impact of one-day changes using a $\sqrt{10}$ multiplier. In FRTB, banks are required to consider changes over periods of 10 days that occurred during a stressed period in the past. Econometricians naturally prefer that non-overlapping periods be used when VaR or ES is being estimated using historical simulation, because they want observations on the losses to be independent. However, this is not feasible when 10-day changes are considered, because it would require a very long historical period. FRTB requires banks to base their estimates on overlapping 10-day periods. The first simulation trial assumes that the percentage changes in all risk factors over the next 10 days will be the same as their changes between Day 0 and Day 10 of the stressed period; the second simulation trial assumes that the percentage changes in all risk factors over the next 10 days will be the same as their changes between Day 1 and Day 11 of the stressed period; and so on.

Banks are first required to calculate ES when 10-day changes are made to all risk factors. (We will denote this by ES_1 .) They are then required to calculate ES when 10-day changes are made to all risk factors in categories 2 and above with risk factors in category 1 being kept constant. (We will denote this by ES_2 .) They are then required to calculate ES when 10-day changes are made to all risk factors in categories 3, 4, and 5 with risk factors in categories 1 and 2 being kept constant. (We will denote this by ES_3 .) They are then required to calculate ES when 10-day changes are made to all risk factors in categories 4 and 5 with risk factors in categories 1, 2, and 3 being kept constant. (We will denote this by ES_4 .) Finally, they are required to calculate ES_5 , which is the effect of making 10-day changes only to category 5 risk factors.

The liquidity-adjusted ES is calculated as

$$\sqrt{ES_1^2 + \sum_{j=2}^5 \left(ES_j \sqrt{\frac{LH_j - LH_{j-1}}{10}} \right)^2} \quad (27.2)$$

where LH_j is the liquidity horizon for category j . To understand equation (27.2), suppose first that all risk factors are in category 1 or 2 so that only ES_1 and ES_2 are calculated. It is assumed that the behavior of all risk factors during a 10-day period is independent of the behavior of category 2 risk factors during a further 10-day period. An extension of the square root rule discussed in Chapter 12 then leads to the liquidity-adjusted ES being

$$\sqrt{ES_1^2 + ES_2^2}$$

Now suppose that there are also category 3 risk factors. The expression $\sqrt{ES_1^2 + ES_2^2}$ would be correct if the category 3 risk factors had a 20-day instead of a 40-day liquidity horizon. We assume that the behavior of the category 3 risk factors over an additional 20 days is independent of the behavior of all the risk factors over the periods already

considered. We also assume that the ES for the category 3 risk factors over 20 days is $\sqrt{2}$ times their ES over 10 days. This leads to a liquidity-adjusted ES of:

$$\sqrt{ES_1^2 + ES_2^2 + 2ES_3^2}$$

Continuing in this way, we obtain equation (27.2). This is referred to as the *cascade approach* to calculating ES (and can be used for VaR as well).

Calculations are carried out for each desk. If there are six desks, this means the internal models approach, as we have described it so far, requires $5 \times 6 = 30$ ES calculations. As mentioned, the use of overlapping time periods is less than ideal because changes in successive historical simulation trials are not independent. This does not bias the results, but it reduces the effective sample size, making results more noisy than they would otherwise be.

FRTB represents a movement away from basing calculations on one-day changes. Presumably the Basel Committee has decided that, in spite of the lack of independence of observations, a measure calculated from 10-day changes provides more relevant information than a measure calculated from one-day changes. This could be the case if changes on successive days are not independent, but changes in successive 10-day periods can reasonably be assumed to be independent.

The calculation of a stressed measure (VaR or ES) requires banks to search for the period in the past when market variable changes would be worst for their current portfolio. (The search must go back as far as 2007.) When Basel II.5 was implemented, a problem was encountered in that banks found that historical data were not available for some of their current risk factors. It was therefore not possible to know how these risk factors would have behaved during the 250-day periods in the past that were candidates for the reference stressed period. FRTB handles this by allowing the search for stressed periods to involve a subset of risk factors, provided that at least 75% of the current risk factors are used. The expected shortfalls that are calculated are scaled up by the ratio of ES for the most recent 12 months using all risk factors to ES for the most recent 12 months using the subset of risk factors. (This potentially doubles the number of ES calculations from 30 to 60.)

Banks are required to calculate ES for the whole portfolio as well for each of six trading desks. The ES for a trading desk is referred to as a *partial expected shortfall*. It is determined by shocking the risk factors belonging to the trading desk while keeping all other risk factors fixed. The sum of the partial expected shortfalls is always greater than the ES for the whole portfolio. What we will refer to as the weighted expected shortfall (WES) is a weighted average of (a) the ES for the whole portfolio and (b) the sum of the partial expected shortfalls. Specifically:

$$WES = \lambda \times EST + (1 - \lambda) \times \sum_j ESP_j$$

where EST is the expected shortfall calculated for the total portfolio and ESP_j is the j th partial expected shortfall. The parameter λ is set by the Basel Committee to be 0.5.

Some risk factors are categorized as *non-modelable*. Specifically, if there are less than 24 observations on a risk factor in a year or more than one month between successive observations, the risk factor is classified as non-modelable. Such risk factors are handled by special rules involving stress tests.

The total capital requirement for day t is

$$\max(\text{WES}_{t-1} + \text{NMC}_{t-1}, m_c \times \text{WES}_{\text{avg}} + \text{NMC}_{\text{avg}})$$

where WES_{t-1} is the WES for day $t - 1$, NMC_{t-1} is the capital charge calculated for non-modelable risk factors on day $t - 1$, WES_{avg} is the average WES for the previous 60 days, and NMC_{avg} is the average capital charge calculated for the non-modelable risk factors over the previous 60 days. The parameter m_c is at minimum 1.5.

27.3.1 Back-Testing

FRTB does not back-test the stressed ES measures that are used to calculate capital under the internal models approach for two reasons. First, it is more difficult to back-test ES than VaR. Second, it is not possible to back-test a stressed measure at all. The stressed data upon which a stressed measure is based are extreme data that statistically speaking are not expected to be observed with the same frequency in the future as they were during the stressed period.

FRTB back-tests a bank's models by asking each trading desk to back-test a VaR measure calculated over a one-day horizon and the most recent 12 months of data. Both 99% and 97.5% confidence levels are to be used. If there are more than 12 exceptions for the 99% VaR or more than 30 exceptions for the 97.5% VaR, the trading desk is required to calculate capital using the standardized approach until neither of these two conditions continues to exist.

Banks may be asked by regulators to carry out other back-tests. Some of these could involve calculating the p -value of the profit or loss on each day. This is the probability of observing a profit that is less than the actual profit or a loss that is greater than the actual loss. If the model is working perfectly, the p -values obtained should be uniformly distributed.

27.3.2 Profit and Loss Attribution

Another test used by the regulators is known as *profit and loss attribution*. Banks are required to compare the actual profit or loss in a day with that predicted by their models. Two measures must be calculated. The measures are:

$$\frac{\text{Mean of } U}{\text{Standard Deviation of } V}$$

$$\frac{\text{Variance of } U}{\text{Variance of } V}$$

where U denotes the difference between the actual and model profit/loss in a day and V denotes the actual profit/loss in a day.⁹ Regulators expect the first measure to be between -10% and $+10\%$ and the second measure to be less than 20% . When there are four or more situations in a 12-month period where the ratios are outside these ranges, the desk must use the standardized approach for determining capital.

27.3.3 Credit Risk

As mentioned, FRTB distinguishes two types of credit risk exposure to a company:

1. *Credit spread risk* is the risk that the company's credit spread will change, causing the mark-to-market value of the instrument to change.
2. *Jump-to-default risk* is the risk that there will be a default by the company.

Under the internal models approach, the credit spread risk is handled in a similar way to market risks. Table 27.1 shows that the liquidity horizon for credit spread varies from 20 to 120 days and the liquidity horizon for a credit spread volatility is 120 days. The jump-to-default risk is handled in the same way as default risks in the banking book. In the internal models approach, the capital charge is based on a VaR calculation with a one-year time horizon and a 99.9% confidence level.

27.3.4 Securitizations

The comprehensive risk measure (CRM) charge was introduced in Basel II.5 to cover the risks in products created by securitizations such as asset-backed securities and collateralized debt obligations (see Section 26.1). The CRM rules allow a bank (with regulatory approval) to use its own models. The Basel Committee has concluded that this is unsatisfactory because there is too much variation in the capital charges calculated by different banks for the same portfolio. It has therefore decided that under FRTB the standardized approach must be used for securitizations.

27.4 Trading Book vs. Banking Book

The FRTB addresses whether instruments should be put in the trading book or the banking book. Roughly speaking, the trading book consists of instruments that the bank intends to trade. The banking book consists of instruments that are expected to be held to maturity. Instruments in the banking book are subject to credit risk capital whereas those in the trading book are subject to market risk capital. The two sorts of capital are calculated in quite different ways. This has in the past given rise to regulatory arbitrage. For example, as we have mentioned in earlier chapters, banks have often chosen to hold credit-dependent instruments in the trading book because they are then subject to less regulatory capital than they would be if they had been placed in the banking book.

⁹The “actual” profit/loss should be the profit and loss that would occur if there had been no trading in a day. This is sometimes referred to as the *hypothetical profit and loss*.

The FRTB attempts to make the distinction between the trading book and the banking book clearer and less subjective. To be in the trading book, it will no longer be sufficient for a bank to have an “intent to trade.” It must be able to trade and manage the underlying risks on a trading desk. The day-to-day changes in value should affect equity and pose risks to solvency. The FRTB provides rules for determining for different types of instruments whether they should be placed in the trading book or the banking book.

An important point is that instruments are assigned to the banking book or the trading book when they are initiated and there are strict rules preventing them from being subsequently moved between the two books. Transfers from one book to another can happen only in extraordinary circumstances. (Examples given of extraordinary circumstances are the closing of trading desks and a change in accounting standards with regard to the recognition of fair value.) Any capital benefit as a result of moving items between the books will be disallowed.

Summary

FRTB is a major change to the way capital is calculated for market risk. After over 20 years of using VaR with a 10-day time horizon and 99% confidence to determine market risk capital, regulators are switching to using ES with a 97.5% confidence level and varying time horizons. The time horizons, which can be as high as 120 days, are designed to incorporate liquidity considerations into the capital calculations. The change that is considered to a risk factor when capital is calculated reflects movements in the risk factor over a period of time equal to the liquidity horizon in stressed market conditions.

The Basel Committee has specified a standardized approach and an internal models approach. Even when they have been approved by their supervisors to use the internal models approach, banks must also implement the standardized approach. Regulatory capital under the standardized approach is based on formulas involving the delta, vega, and gamma exposures of the trading book. Regulatory capital under the internal models approach is based on the calculation of stressed expected shortfall. Calculations are carried out separately for each trading desk.

Further Reading

Bank for International Settlements. “Minimum Capital Requirements for Market Risk,” January 2019.

Practice Questions and Problems (Answers at End of Book)

- 27.1 Outline the differences between the way market risk capital is calculated in (a) Basel I, (b) Basel II.5, and (c) the FRTB.
- 27.2 Use footnote 4 to verify that when losses have a normal distribution with mean μ and standard deviation σ the 97.5% expected shortfall is $\mu + 2.338\sigma$.

- 27.3 Explain why the use of overlapping time periods proposed by the FRTB does not give independent observations on the changes in variables.
- 27.4 What are the advantages of expected shortfall over value at risk?
- 27.5 What is the difference between the trading book and the banking book? Why are regulators concerned about specifying whether an instrument should be one or the other in the FRTB?
- 27.6 How are credit trades handled under the FRTB?

Further Question

- 27.7 Suppose that an investor owns the \$10 million portfolio in Table 12.1 on July 8, 2020. Suppose that the 250 days ending April 30, 2020, constitute the stressed period for the portfolio. Calculate the 97.5% expected shortfall using the overlapping periods method in conjunction with historical simulation and the cascade approach. Relevant data on the indices is on the author's website (see "Worksheets for Value at Risk Example"). For the purposes of this problem, assume that S&P 500 and FTSE have a 10-day liquidity horizon, CAC 40 has a 40-day liquidity horizon, and Nikkei 225 has a 20-day liquidity horizon. For each day during the stressed period, consider the change in a variable over a 10-day period ending on the day.

Chapter 28

Economic Capital and RAROC

E*conomic capital* (sometimes referred to as *risk capital*) is a financial institution's own internal estimate of the capital it needs for the risks it is taking. It is different from regulatory capital, which tends to be based on one-size-fits-all rules. Economic capital can be regarded as a “currency” for risk-taking within a financial institution. A business unit can take a certain risk only when it is allocated the appropriate economic capital for that risk. The profitability of a business unit is measured relative to the economic capital allocated to the unit.

In this chapter, we discuss the approaches a financial institution uses to arrive at estimates of economic capital for particular risk types and particular business units, and how these estimates are aggregated to produce a single economic capital estimate for the whole financial institution. We also discuss risk-adjusted return on capital, or RAROC. This is the return earned by a business unit on the capital assigned to it. RAROC can be used to assess the past performance of business units. It can also be used to forecast future performance of the units and decide on the most appropriate way of allocating

capital in the future. It provides a basis for determining whether some activities should be discontinued and others expanded.

28.1 Definition of Economic Capital

Economic capital is usually defined as the amount of capital a financial institution needs in order to absorb losses over one year with a certain confidence level. The confidence level is therefore the probability that the bank's losses will not exceed the economic capital in one year. The confidence level depends on the financial institution's objectives. A common objective for a large international bank is to maintain a AA credit rating. Corporations rated AA have a one-year probability of default of about 0.02%. This suggests that the confidence level should be set as high as 99.98% for economic capital to be a guide as to what is necessary to maintain a AA rating. The probability that capital will be insufficient to absorb losses so that a default results is then the required 0.02%. For a bank wanting to maintain a BBB credit rating, the confidence level can be lower. A BBB-rated corporation has a probability of about 0.2% of defaulting in one year, so that a confidence level of 99.8% gives the capital necessary to maintain a BBB rating.

Capital is required to cover unexpected loss. This is defined as the difference between the actual loss and the expected loss. As explained in Chapter 25, the idea here is that expected losses should be taken account of in the way a financial institution prices its products so that only unexpected losses require capital. As indicated in Figure 28.1, the economic capital is the difference between expected loss and the X percentile point on the probability distribution of the loss, where $X\%$ is the confidence level.

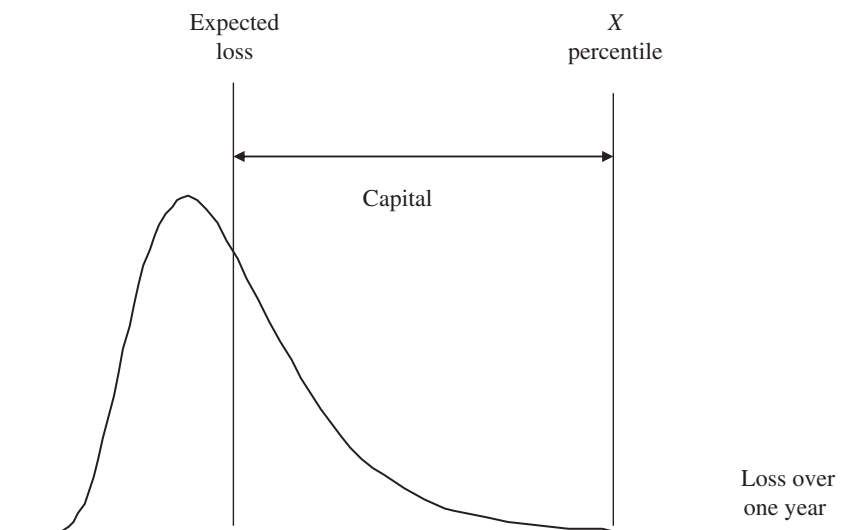


Figure 28.1 Calculation of Economic Capital from One-Year Loss Distribution
 $X\%$ is the confidence level.

Example 28.1

Suppose that the main business of an AA-rated bank is lending in a certain region of the world. The bank estimates its losses as 1% of outstanding loan principal per year on average. The one-year 99.98% VaR is estimated as 5% of the outstanding loan principal. The economic capital required per \$100 of loans made is therefore \$4. (This is the difference between the one year 99.98% VaR and the expected loss.)

28.1.1 Approaches to Measurement

There are two broad approaches to measuring economic capital: the “top-down” and “bottom-up” approaches. In the top-down approach, the volatility of the financial institution’s assets is estimated and then used to calculate the probability that the value of the assets will fall below the value of the liabilities by the end of the time horizon. A theoretical framework that can be used for the top-down approach is Merton’s model, which was discussed in Section 17.8.

The approach most often used is the bottom-up approach where loss distributions are estimated for different types of risk and different business units and then aggregated. The first step in the aggregation can be to calculate probability distributions for losses by risk type or losses by business unit. A final aggregation gives a probability distribution of total losses for the whole financial institution.

The various risks facing a bank or other financial institution are summarized in Figure 28.2. As we saw in Chapter 20, bank regulators have chosen to define operational risk as: “The risk of loss resulting from inadequate or failed internal processes, people, and systems or from external events.” Operational risk includes model risk and legal risk, but it does not include risk arising from strategic decisions and reputational risk. We will collectively refer to the latter risks as *business risk*. Regulatory capital is not required for business risk, but some banks do assess economic capital for it.

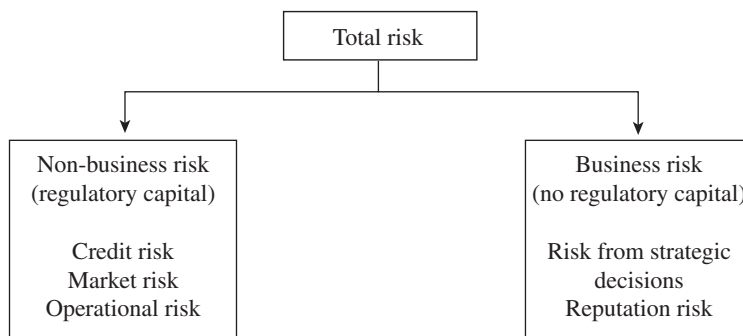


Figure 28.2 Categorization of Risks Faced by a Bank and Regulatory Capital Requirements

28.2 Components of Economic Capital

In earlier chapters, we covered approaches used to calculate loss distributions for different types of risk. This section reviews some of the key issues.

28.2.1 *Market Risk Economic Capital*

In Chapters 12 and 13, we discussed the historical simulation and model-building approaches for estimating the probability distribution of losses or gains from market risk. As explained, this distribution is usually calculated in the first instance with a one-day time horizon. Regulatory capital for market risk in Basel I and Basel II.5 is calculated from 10-day 99% VaR estimates. In FRTB, as described in Chapter 27, ES is used with time horizons determined by liquidity.

When calculating economic capital, we usually want to use the same time horizon and confidence level for all risks. The time horizon is usually one year and, as already discussed, the confidence level can be as high as 99.98%. Simple assumptions are (a) that the probability distributions of gains and losses for each day during the year are the same and (b) that the distributions are independent. The one-year loss or gain distribution is then approximately normal. Assuming 252 business days in the year, the standard deviation of the one-year loss or gain equals the standard deviation of the daily loss or gain multiplied by $\sqrt{252}$. If the average loss is zero, the 99.98% VaR is then 3.54 times the standard deviation of the one-year loss or gain. The 99.8% VaR is 2.88 times the standard deviation of the one-year loss or gain. Note that we are not assuming that the daily losses or gains are normal. All we are assuming is that they are independent and identically distributed. The central limit theorem of statistics tells us that the sum of many independent identically distributed variables is approximately normal.

Example 28.2

Suppose that the one-day standard deviation of market risk losses or gains for a bank is \$5 million. The one-year 99.8% VaR is $2.88 \times \sqrt{252} \times 5 = 228.6$ or \$228.6 million.

In practice, the losses (gains) on different days are not perfectly independent and identically distributed. The independence assumption can be relaxed by assuming a constant autocorrelation between the returns on successive days. This allows the result in equation (11.5) to be used and, providing the autocorrelation is not too large, the loss (gain) in a year can still be assumed to be approximately normal. Sometimes, it may be appropriate to develop a more elaborate model of how the standard deviation of the daily loss (gain) evolves through time. Monte Carlo simulation can then be used to aggregate the daily distributions and the result may not then be a normal distribution.

28.2.2 Credit Risk Economic Capital

Although Basel II gives banks that use the internal-ratings-based approach for regulatory capital a great deal of freedom, it does not allow them to choose their own credit correlation model and correlation parameters. When calculating economic capital, banks are free to make the assumptions they consider most appropriate for their situation. CreditMetrics, which is explained in Section 19.4, is a popular approach for calculating credit risk economic capital.

Another approach that has been proposed is Credit Risk Plus, which is described in Section 19.3. This approach borrows a number of ideas from actuarial science to calculate a probability distribution for losses from defaults. Whereas CreditMetrics calculates the loss from downgrades and defaults, Credit Risk Plus calculates losses from defaults only.

In calculating credit risk economic capital, a financial institution can choose to adopt a conditional or unconditional model. In a conditional (cycle-specific) model, the expected and unexpected losses take account of current economic conditions. In an unconditional (cycle-neutral) model, they are calculated by assuming economic conditions that are in some sense an average of those experienced through the cycle. Rating agencies aim to produce ratings that are unconditional. In addition, when regulatory capital is calculated using the internal-ratings-based approach, the PD and LGD estimates should be unconditional. Obviously, it is important to be consistent when economic capital is calculated. If expected losses are conditional, unexpected losses should also be conditional. If expected losses are unconditional, the same should be true of unexpected losses.

Whatever the approach used, a Monte Carlo simulation is usually necessary to calculate the probability distribution of credit losses. As explained in Chapters 20 and 21, derivatives complicate the calculations because of the uncertainty about the exposure at the time of a default or downgrade.

28.2.3 Operational Risk Economic Capital

The procedure used by bank regulators for operational risk is described in Section 20.4. Many banks are likely to choose this approach for economic capital. But some may attempt to use more sophisticated models that lead to a complete distribution of losses being estimated.

28.2.4 Business Risk Economic Capital

As mentioned earlier, business risk includes strategic risk (relating to a bank's decision to enter new markets and develop new products) and reputational risk. Business risk is even more difficult to quantify than operational risk and estimates are likely to be largely subjective. However, attempting to quantify business risk can be a useful exercise, as it can lead to risk managers within a financial institution acquiring a good understanding of the portfolio of business risks being taken. This enables them to assess the marginal impact on total risk of new strategic initiatives that are being contemplated.

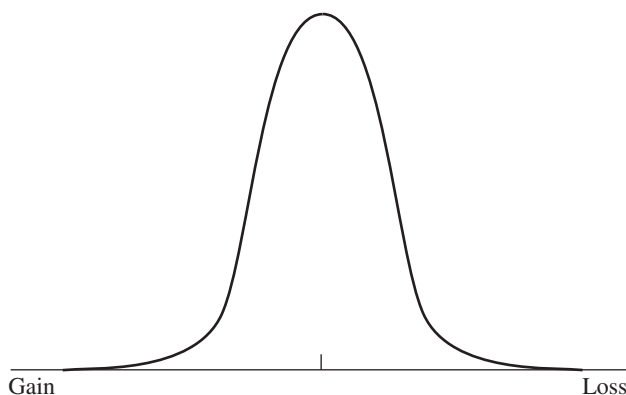


Figure 28.3 Loss Probability Density Function for Market Risk

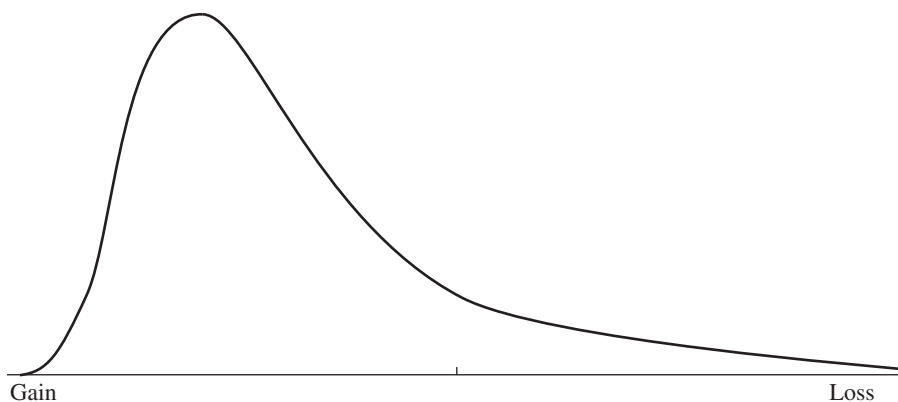


Figure 28.4 Loss Probability Density Function for Credit Risk

28.3 Shapes of the Loss Distributions

The loss probability distributions for market, credit, and operational risk are very different. Rosenberg and Schuermann (2004) used data from a variety of different sources to estimate typical shapes for these distributions.¹ These are shown in Figures 28.3 to 28.5. The market risk loss distribution (see Figure 28.3) is symmetrical but not perfectly normally distributed. A t -distribution with 11 degrees of freedom has been shown to provide a good fit. The credit risk loss distribution in Figure 28.4 is quite skewed, as one would expect. The operational risk distribution in Figure 28.5 has a quite extreme shape. Most of the time, losses are modest, but occasionally they are very large.

We can characterize a distribution by its second, third, and fourth moments. Loosely speaking, the second moment measures standard deviation (or variance), the

¹See J. V. Rosenberg and T. Schuermann, "A General Approach to Integrated Risk Management with Skewed, Fat-Tailed Risks," Federal Reserve Bank of New York, Staff Report no. 185, May 2004.

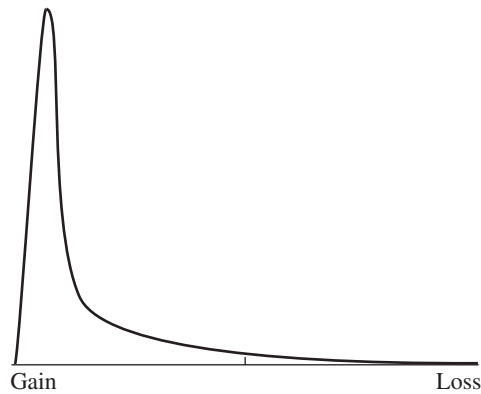


Figure 28.5 Loss Probability Density Function for Operational Risk

Table 28.1 Characteristics of Loss Distributions for Different Risk Types

	Second Moment (standard deviation)	Third Moment (skewness)	Fourth Moment (kurtosis)
Market risk	High	Zero	Low
Credit risk	Moderate	Moderate	Moderate
Operational risk	Low	High	High

third moment measures skewness, and the fourth moment measures kurtosis (i.e., the heaviness of tails). Table 28.1 summarizes the properties of typical loss distributions for market, credit, and operational risk.

28.4 Relative Importance of Risks

The relative importance of different types of risk depends on the business mix. Credit risk is important in commercial lending, retail lending, and a financial institution's derivatives business. Market risk is important in trading and some investment banking activities. But as mentioned in Chapter 20, operational risk (particularly cyber risk, legal risk, and compliance risk) is now considered by many observers to be the most important risk for financial institutions.

Operational risk has always been considered to be the most important risk in asset management. If the asset manager is found to be negligent in some way, there are liable to be expensive investor lawsuits. Business Snapshot 28.1 gives one example of this. Another high-profile example is provided by Unilever's pension plan. Mercury Asset Management, owned by Merrill Lynch, pledged not to underperform a benchmark index by more than 3%. Between January 1997 and March 1998, it underperformed the index by 10.5%. Unilever sued Merrill Lynch for \$185 million and the matter was settled out of court. The Spanish bank Santander incurred operational risk losses in 2009 on funds it managed for investors and placed with Bernard Madoff, who, it transpired, was running a \$50 billion Ponzi scheme.

BUSINESS SNAPSHOT 28.1**The EGT Fund**

In 1996, Peter Young was fund manager at Deutsche Morgan Grenfell, a subsidiary of Deutsche Bank. He was responsible for managing a fund called the European Growth Trust (EGT). It had grown to be a very large fund and Young had responsibilities for managing over one billion pounds of investors' money.

Certain rules applied to EGT. One of these was that no more than 10% of the fund could be invested in unlisted securities. Peter Young violated this rule in a way that it can be argued benefited him personally. When the facts were uncovered, he was fired and Deutsche Bank had to compensate investors. The total cost to Deutsche Bank was over 200 million pounds.

28.4.1 Interactions between Risks

There are interactions between the different types of risks. If a financial institution's counterparty defaults, credit risk exists only if market variables have moved so that the uncollateralized value of the derivative to the financial institution is positive. Another interaction is that the probability of default by a counterparty may depend on the value of a financial institution's contract (or contracts) with the counterparty. This is the wrong-way issue discussed in Chapter 18. If the counterparty has entered into the contract for hedging purposes, the dependence should be small. However, if the contract has been entered into for speculative purposes and the contract is large in relation to the size of the counterparty, the dependence is likely to be important.

As the Long-Term Capital Management saga clearly shows, there can be interactions between liquidity risks and market risks (see Business Snapshot 16.1). There are also interactions between operational risks and market risks. It is unlikely that we would know about the activities of Jérôme Kerviel at Société Générale if he had guessed right about market movements (see Business Snapshot 5.5). Similarly, we are unlikely to hear about a violation of the rules for a fund (such as the one in Business Snapshot 28.1) if the violation leads to a gain rather than a loss.

28.5 Aggregating Economic Capital

A financial institution typically calculates market, credit, operational, and (possibly) business risk loss distributions for a number of different business units. It is then faced with the problem of aggregating the loss distributions to calculate a total economic capital for the whole enterprise.

The simplest approach is to assume that the total economic capital for a set of n different risks is the sum of the economic capital amounts for each risk considered separately, so that

$$E_{\text{total}} = \sum_{i=1}^n E_i \quad (28.1)$$

where E_{total} is the total economic capital for the financial institution facing n different risks and E_i is the economic capital for the i th risk considered on its own. This is what the Basel Committee does for regulatory capital. The total regulatory capital a bank is required to keep is the sum of the regulatory capital amounts for credit, market, and operational risks.

Equation (28.1) is clearly a very conservative assumption. It assumes perfect correlation. In the context of economic capital calculations where the confidence level is, say, 99.9%, it would mean that, if the market risk loss for a financial institution is in the extreme 0.1% tail of the distribution, the same is true for the credit risk loss and the operational risk loss. Rosenberg and Schuermann estimate the correlation between market risk and credit risk to be approximately 50% and the correlation between each of these risks and operational risk to be approximately 20%. They estimate that equation (28.1) when used as a way of aggregating market, credit, and operational risks overstates the total capital required by about 40%.

28.5.1 Assuming Normal Distributions

A simple assumption when aggregating loss distributions is that they are normally distributed. The standard deviation of the total loss from n sources of risk is then

$$\sigma_{\text{total}} = \sqrt{\sum_{i=1}^n \sum_{j=1}^n \sigma_i \sigma_j \rho_{ij}} \quad (28.2)$$

where σ_i is the standard deviation of the loss from the i th source of risk and ρ_{ij} is the correlation between risk i and risk j . The capital requirement can be calculated from this. For example, the excess of the 99.9% VaR over the expected loss is 3.09 times the number calculated in equation (28.2).

This approach tends to underestimate the capital requirement because it takes no account of the skewness and kurtosis of the loss distributions. Rosenberg and Schuermann estimate that, when the approach is applied to aggregating market, credit, and operational risks, the total capital is underestimated by about 40%.

28.5.2 Using Copulas

A more sophisticated approach to aggregating loss distributions is by using copulas. Copulas were discussed in Chapter 9. Each loss distribution is mapped on a percentile to percentile basis to a standard well-behaved distribution. A correlation structure between

the standard distributions is defined and this indirectly defines a correlation structure between the original distributions.

Many different copulas can be defined. The simplest is the Gaussian copula, where the standard distributions are assumed to be multivariate normal. Other alternatives, with more tail correlation, are explained in Section 9.5.

28.5.3 The Hybrid Approach

The approach in Section 11.9 for aggregating VaR is a popular approach for aggregating economic capital estimates. It is sometimes referred to as the *hybrid approach*. It involves calculating the economic capital for a portfolio of risks from the economic capital for the individual risks using

$$E_{\text{total}} = \sqrt{\sum_{i=1}^n \sum_{j=1}^n E_i E_j \rho_{ij}} \quad (28.3)$$

When the distributions are normal, this approach is exactly correct. When distributions are non-normal, the hybrid approach gives an approximate answer—but one that reflects any heaviness in the tails of the individual loss distributions. Rosenberg and Schuermann find that the answers given by the hybrid approach are reasonably close to those given by copula models.

Example 28.3

Suppose that the estimates for economic capital for market, credit, and operational risk for two business units are as shown in Table 28.2. The correlations between the losses are shown in Table 28.3. The correlation between credit risk and market risk within the same business unit is 0.5, and the correlation between operational risk and either credit or market risk within the same business unit is 0.2. (These correspond to the estimates of Rosenberg and Schuermann mentioned earlier.) The correlation between two different risk types in two different business units is zero. The correlation between market risks across business units is 0.4. The correlation between credit risks across business units is 0.6. The correlation between operational risks across business units is zero.

Table 28.2 Economic Capital Estimates for Example 28.3

	Business Units	
	1	2
Market Risk	30	40
Credit Risk	70	80
Operational Risk	30	90

Table 28.3 Correlations between Losses in Example 28.3

	MR-1	CR-1	OR-1	MR-2	CR-2	OR-2
MR-1	1.0	0.5	0.2	0.4	0.0	0.0
CR-1	0.5	1.0	0.2	0.0	0.6	0.0
OR-1	0.2	0.2	1.0	0.0	0.0	0.0
MR-2	0.4	0.0	0.0	1.0	0.5	0.2
CR-2	0.0	0.6	0.0	0.5	1.0	0.2
OR-2	0.0	0.0	0.0	0.2	0.2	1.0

MR, CR, and OR refer to market risk, credit risk, and operational risk; 1 and 2 refer to business units.

We can aggregate the economic capital using the hybrid approach. The total market risk economic capital is

$$\sqrt{30^2 + 40^2 + 2 \times 0.4 \times 30 \times 40} = 58.8$$

The total credit risk economic capital is

$$\sqrt{70^2 + 80^2 + 2 \times 0.6 \times 70 \times 80} = 134.2$$

The total operational risk economic capital is

$$\sqrt{30^2 + 90^2} = 94.9$$

The total economic capital for business unit 1 is

$$\sqrt{30^2 + 70^2 + 30^2 + 2 \times 0.5 \times 30 \times 70 + 2 \times 0.2 \times 30 \times 30 + 2 \times 0.2 \times 70 \times 30} = 100.0$$

The total economic capital for business unit 2 is

$$\sqrt{40^2 + 80^2 + 90^2 + 2 \times 0.5 \times 40 \times 80 + 2 \times 0.2 \times 40 \times 90 + 2 \times 0.2 \times 80 \times 90} = 153.7$$

The total enterprise-wide economic capital is the square root of

$$\begin{aligned} &30^2 + 40^2 + 70^2 + 80^2 + 30^2 + 90^2 + 2 \times 0.4 \times 30 \times 40 + 2 \times 0.5 \times 30 \times 70 \\ &+ 2 \times 0.2 \times 30 \times 30 + 2 \times 0.5 \times 40 \times 80 + 2 \times 0.2 \times 40 \times 90 \\ &+ 2 \times 0.6 \times 70 \times 80 + 2 \times 0.2 \times 70 \times 30 + 2 \times 0.2 \times 80 \times 90 \end{aligned}$$

or 203.2.

There are clear diversification benefits. The sum of the economic capital estimates for market, credit, and operational risk is $58.8 + 134.2 + 94.9 = 287.9$ and the sum of the economic capital estimates for two business units is $100 + 153.7 = 253.7$. Both of these are greater than the total economic capital estimate of 203.2.

28.6 Allocation of Economic Capital

Suppose that the sum of the economic capital for the business units in a bank, $\sum_{i=1}^n E_i$, is \$2 billion and that the total economic capital for the whole bank, after taking less-than-perfect correlations into account, is \$1.3 billion (= 65% of the sum of the E_i). The \$0.7 billion is a diversification gain to the bank. How should the total economic capital be allocated to the business units?

A simple approach is to allocate $0.65E_i$ to business unit i . However, this is probably not the best approach. Consider a situation where there are 50 business units and that two particular units both have an economic capital of \$100 million. Suppose that, when the first business unit is excluded from the calculations, the bank's economic capital decreases by \$60 million and that, when the second business unit is excluded from the calculation, the bank's economic capital decreases by \$10 million. Arguably, the first business unit should have more economic capital than the second, because its incremental impact on the bank's total economic capital is greater.

The issues here are analogous to those we discussed in Section 11.8 for allocating VaR. One approach is to calculate incremental economic capital for each business unit and then allocate economic capital to business units in proportion to their incremental capital. (Incremental capital is the difference between the total economic capital with and without the business unit.) A popular alternative is to work with the component economic capital. This involves allocating

$$x_i \frac{\partial E_{\text{total}}}{\partial x_i}$$

to the i th business unit, where x_i is the investment in the i th business unit. As we pointed out in Section 11.8, a result known as Euler's theorem then ensures that the total of the allocated capital is the total economic capital E_{total} .

Define Q_i as the increase in the total economic capital when we increase x_i by a small amount Δx_i . A discrete approximation for the component economic capital for business unit i is

$$\frac{Q_i}{y_i} \quad (28.4)$$

where $y_i = \Delta x_i / x_i$.

Example 28.4

Consider again Example 28.3. The total economic capital is 203.2. The economic capital calculated for Business Unit 1 is 100 and that calculated for Business Unit 2 is 153.7.

A simple approach would allocate $100/253.7$ of the total economic capital to Business Unit 1 and $153.7/253.7$ of the economic capital to Business Unit 2. This would result in 80.1 for Business Unit 1 and 123.1 to Business Unit 2.

The incremental effect of Business Unit 1 on the total economic capital is $203.2 - 153.7 = 49.5$. Similarly, the incremental effect of Business Unit 2 on the total economic

capital is $203.2 - 100 = 103.2$. The two incremental capitals do not add up to the total capital (as is usually the case). However, we could use them as a basis for allocating the total capital. We would then allocate $49.5/(49.5 + 103.2)$ of the capital to Business Unit 1 and $103.2/(49.5 + 103.2)$ of it to Business Unit 2. This would result in 65.9 for Business Unit 1 and 137.3 for Business Unit 2.

To apply equation (28.4) and allocate component economic capital to each unit, we could calculate the partial derivative analytically. Alternatively, we can use a numerical calculation. When we increase the size of Business Unit 1 by 1%, its economic capital amounts for market, credit, and operational risk in Table 28.2 increase to 30.3, 70.7, and 30.3, respectively. The total economic capital becomes 203.906 so that $Q_1 = 203.906 - 203.224 = 0.682$.

When we increase the size of Business Unit 2 by 1%, its economic capital amounts for market, credit, and operational risk in Table 28.2 increase to 40.4, 80.8, and 90.9, respectively. The total economic capital becomes 204.577 so that $Q_2 = 204.577 - 203.224 = 1.353$.

In this case, because we are considering 1% increases in the size of each unit, $\gamma_1 = \gamma_2 = 0.01$. From equation (28.4), the economic capital allocations to the two business units are 68.2 and 135.3. (These do not quite add up to the total economic capital of 203.2 because we approximated the partial derivative.)

28.7 Deutsche Bank's Economic Capital

Deutsche Bank publishes the results of its economic capital calculations in its annual financial statements. Table 28.4 summarizes the economic capital and regulatory capital reported at the end of 2021. Capital is calculated for credit risk, market risk, operational risk, and strategic (i.e., business) risk. Deutsche Bank calculated a diversification benefit reflecting the lack of perfect correlation between the different risk types. The total economic capital was about 23.5 billion euros.

The actual capital held was about 46.5 billion euros of core Tier 1 capital (i.e., common equity), 8.3 billion euros of additional Tier 1 capital, and about 7.3 billion euros of Tier 2 capital. Table 28.4 shows capital amounts as a percentage of risk-weighted assets. Deutsche Bank exceeds the capital requirements in Basel III, including the G-SIB additional capital, which in the case of Deutsche Bank was 1.5% of risk-weighted assets.

Table 28.5 summarizes how this capital was allocated to business units. A more detailed analysis given by Deutsche Bank shows that credit risk accounted for about 70% of the corporate bank's economic capital. Credit risk was also the most important component for the investment bank, but market risk and operational risk were relatively more important for the investment bank than for the corporate bank.

Table 28.4 Deutsche Bank's Economic Capital and Regulatory Capital, December 2021 (millions of euros)

Credit risk	11,725
Market risk	7,920
Operational risk	4,937
Strategic risk	3,173
Diversification benefit	(4,213)
Total economic capital	23,542
Total risk-weighted assets	351,629
Common Equity Tier 1 capital (% of risk-weighted assets)	13.2%
All Tier 1 capital (% of risk-weighted assets)	15.6%
Total (Tier 1 plus Tier 2) capital (% of risk-weighted assets)	17.7%

Table 28.5 Allocation of Deutsche Bank's Economic Capital to Divisions, December 2021 (millions of euros)

Division	Allocated Economic Capital
Corporate Bank	3,500
Investment Bank	7,442
Private Bank	3,183
Asset Management	267
Capital Release Unit	1,309
Corporate & Other	7,840
Total	23,542

28.8 RAROC

Risk-adjusted performance measurement (RAPM) has become an important part of how business units are assessed. There are many different approaches, but all have one thing in common. They compare return with capital employed in a way that incorporates an adjustment for risk.

The most common approach is to compare expected return with economic capital. This is usually referred to as RAROC (risk-adjusted return on capital). The formula is

$$\text{RAROC} = \frac{\text{Revenues} - \text{Costs} - \text{Expected losses}}{\text{Economic capital}} \quad (28.5)$$

The numerator may be calculated on a pre-tax or post-tax basis. Sometimes, a risk-free rate of return on the economic capital is calculated and added to the numerator.

Example 28.5

A bank rated AA estimates its losses on loans as 1% of the outstanding principal on average. The 99.98% VaR (i.e., the loss exceeded only 0.02% of the time) is 5% of

outstanding loans. As indicated in Example 28.1, the economic capital required per \$100 of loans is \$4, which is the difference between the 99.98% VaR and the expected loss. (This ignores diversification benefits that would in practice be allocated to the lending unit as a result of less than perfect correlation with the bank's other activities.) The spread between the cost of funds and the interest charged is 2.5%. Subtracting from this the expected loan loss of 1%, the expected contribution per \$100 of loans is \$1.50. Assuming that the lending unit's administrative costs total 0.7% of the amount loaned, the expected profit is reduced to \$0.80 per \$100 in the loan portfolio. RAROC is therefore

$$\frac{0.80}{4} = 20\%$$

An alternative calculation would add the interest on the economic capital to the numerator. Suppose that the risk-free interest rate is 2%. Then $0.02 \times 4 = 0.08$ is added to the numerator so that RAROC becomes

$$\frac{0.88}{4} = 22\%$$

As pointed out by Matten (2000), it is more accurate to refer to the approach in equation (28.5) as RORAC (return on risk-adjusted capital) rather than RAROC, because it is the capital (i.e., the denominator) not the return that reflects risk.² In theory, RAROC should involve adjusting the return (i.e., the numerator) for risk. In practice, RAROC is used to denote a wide range of return on capital calculations.

RAROC can be calculated ex-ante (before the start of the year) or ex-post (after the end of the year). Ex-ante calculations are based on estimates of expected profit. Ex-post calculations are based on actual profit results. Ex-ante calculations are typically used to decide whether a particular business unit should be expanded or contracted. Ex-post calculations are typically used for performance evaluation and bonus calculations.

It is usually not appropriate to base a decision to expand or contract a particular business unit on an ex-post analysis (although there is a natural temptation to do this). It may be that results were bad for the most recent year because of a one-time event (e.g., an unusually high credit loss or a particularly high operational risk loss). Key strategic decisions should be based on expected long-term results.

Summary

Economic capital is the capital that a bank or other financial institution deems necessary for the risks it is bearing. When calculating economic capital, a financial institution is free to adopt any approach it likes. It does not have to use the one proposed by regulators. Typically, it estimates economic capital for credit risk, market risk, operational

²See C. Matten, *Managing Bank Capital: Capital Allocation and Performance Measurement*, 2nd ed. (Chichester, UK: John Wiley & Sons, 2000).

risk, and (possibly) business risk for its business units and then aggregates the estimates to produce an estimate of the economic capital for the whole enterprise. The risks are usually assumed to be less than perfectly correlated. The benefits of diversification are estimated and allocated to business units. Usually the approach used is designed to reflect the incremental impact of the business units on the total economic capital.

The one-year loss distributions for market risk, credit risk, and operational risk are quite different. The loss distribution for market risk is symmetrical. For credit risk it is skewed, and for operational risk it is highly skewed with very heavy tails.

Financial institutions calculate RAROC, the risk adjusted return on economic capital, for each business unit by dividing the profit of the business unit by its allocated economic capital. The expected future RAROC is used to decide which areas of the business should be expanded or contracted. The actual RAROC achieved by a business unit is used for performance evaluation.

Further Reading

Dev, A. *Economic Capital: A Practitioner's Guide*. London: Risk Books, 2004.

Matten, C. *Managing Bank Capital: Capital Allocation and Performance Measurement*. 2nd ed. Chichester, UK: John Wiley & Sons, 2000.

Rosenberg, J. V., and T. Schuermann. "A General Approach to Integrated Risk Management with Skewed, Fat-Tailed Risks." Federal Reserve Bank of New York, Staff Report no. 185, May 2004.

Practice Questions and Problems (Answers at End of Book)

- 28.1 What is the difference between economic capital and regulatory capital?
- 28.2 What determines the confidence level used by a AA-rated financial institution in its economic capital calculations?
- 28.3 What types of risk are included in business risk?
- 28.4 In what respects are the models used to calculate economic capital for market risk, credit risk, and operational risk likely to be different from those used to calculate regulatory capital?
- 28.5 Suppose that the credit loss in a year has a lognormal distribution. The logarithm of the loss is normal with mean 0.5 and standard deviation 4. What is the economic capital requirement if a confidence level of 99.97% is used?
- 28.6 Suppose that the economic capital estimates for two business units are as follows:

	Business Units	
	1	2
Market Risk	20	40
Credit Risk	40	30
Operational Risk	70	10

The correlations are as in Table 28.3. Calculate the total economic capital for each business unit and the two business units together.

- 28.7 In Problem 28.6, what is the incremental effect of each business unit on the total economic capital? Use this to allocate economic capital to business units. What is the impact on the economic capital of each business unit increasing by 0.5%? Show that your results are consistent with Euler's theorem.
- 28.8 A bank is considering expanding its asset management operations. The main risk is operational risk. It estimates the expected operational risk loss from the new venture in one year to be \$2 million and the 99.97% VaR (arising from a small chance of an large investor lawsuit) to be \$40 million. The expected fees it will receive from investors for the funds under administration are \$12 million per year, and administrative costs are expected to be \$3 million per year. Estimate the before-tax RAROC.
- 28.9 RAROC can be used in two different ways. What are they?

Further Questions

- 28.10 Suppose that daily gains (losses) are normally distributed with a standard deviation of \$5 million.
- (a) Estimate the minimum regulatory capital the bank is required to hold. (Assume a multiplicative factor of 4.0.)
 - (b) Estimate the economic capital using a one-year time horizon and a 99.9% confidence level assuming that there is a correlation of 0.05 between gains (losses) on successive days.
- 28.11 Suppose that the economic capital estimates for two business units are

	Business Units	
	1	2
Market Risk	10	50
Credit Risk	30	30
Operational Risk	50	10

The correlation between market risk and credit risk in the same business unit is 0.3. The correlation between credit risk in one business unit and credit risk in another is 0.7. The correlation between market risk in one business unit and market risk in the other is 0.2. All other correlations are zero. Calculate the total economic capital. How much should be allocated to each business unit?

- 28.12 Suppose that a bank's sole business is to lend in two regions of the world. The lending in each region has the same characteristics as in Example 28.5 of Section 28.8. Lending to Region A is three times as great as lending to Region B. The correlation between loan losses in the two regions is 0.4. Estimate the total RAROC.

- 28.13 Suppose daily losses (gains) from trading are independent and normally distributed with mean zero. Calculate in terms of the standard deviation of the daily losses (gains) (a) the basic Basel I regulatory capital requirement calculated as 3 times the 10-day 99% VaR and (b) the economic capital calculated using a 99.97% confidence level and a one-year time horizon. Would you expect the economic and regulatory capital to become closer together or farther apart if daily losses/gains are generated by a distribution with much heavier tails than the normal distribution? What would you expect to be the impact of the daily losses/gains exhibiting positive autocorrelation?

Part Seven

OTHER TOPICS

Chapter 29

Financial Innovation

In the 1980s, when booking a flight from New York to London, a customer would contact a travel agent, who would check availability, present options, receive instructions, and then finally make a reservation. Today, that customer is likely to go to the airline's website and make the booking directly with the airline. What has happened is a process known as *disintermediation*. The travel agent as an intermediary is no longer needed.

This does not mean that there is no need at all for intermediaries in the travel business. Online services such as Expedia and Travelocity have sprung up to be of assistance when customers do not know which airline to use or want to quickly compare prices offered by different airlines. But the nature of the intermediaries in the travel business has changed, and the human involvement in the process of making straightforward flight arrangements has largely disappeared. The creation of new technology-based intermediaries like Expedia and Travelocity is referred to as *reintermediation*. Disintermediation followed by reintermediation is a common pattern in technological change.

Banks and other financial companies are intermediaries, like the travel agents of the 1980s, and they are similarly in danger of having the services they provide disrupted by technology. A commercial bank is an intermediary that channels money from depositors to borrowers. It also is an intermediary in the payments system, clearing checks and transferring funds as instructed by clients. Investment banks are intermediaries that facilitate the issue of stocks and bonds. Banks and other financial companies are intermediaries when they perform wealth management services for clients, channeling their savings to appropriate investment products.

FinTech is concerned with the application of technology to financial services. Many applications of technology such as ATMs, online banking, and online trading have already taken place. This chapter looks at how financial services are likely to be impacted by FinTech in the future, which activities are most likely to be digitized, how customers will respond, and how financial services companies can restructure themselves to adjust to rapid technological changes. These are important questions for today's risk management professionals.

29.1 Technological Advances

The increase in computer processing power in the first 50 years of the computer age has been truly astounding, and this has made possible many societal advances. In 1965 Gordon Moore, co-founder of Intel, correctly predicted that computer processing power would double every two years. A single smartphone now has more processing power than NASA had when it landed a man on the moon in 1969!

Many activities formerly carried out by humans have been automated. An interesting development is *robotics process automation* (RPA). This is a software application that replicates the actions of human beings when they interact with a company's systems. It is an alternative to automating by changing the systems themselves. The software robot is a virtual worker that can be trained to perform certain tasks in much the same way as a human would be trained. The tasks could for example be concerned with opening a new customer account or hiring a new employee. Typically, data are sent to a number of different systems within the organization.

29.1.1 Machine Learning

The first industrial revolution, which started in the 18th century, was concerned with using steam and water power. The second industrial revolution, which started in the 19th century, was concerned with electricity and mass production. The third industrial revolution, which started in the mid-20th century, was concerned with digitization: the automation of routine tasks previously carried out by humans. The fourth industrial revolution, which we are now living through, is concerned with creating intelligence from data.

Machine learning is at the center of the fourth industrial revolution.¹ The difference between the third and fourth industrial revolutions can be illustrated by considering the use of computers by a bank to handle loan applications. The loan officer knows certain variables (referred to in machine learning as *features*) describing loan applicants (the credit score of the applicant, the applicant's income, other outstanding debt of the applicant, etc.) If the way these features are used to determine whether an application is accepted

¹See John Hull, *Machine Learning in Business: An Introduction to the World of Data Science*. 3rd edition, 2021, www.rotman.utoronto.ca/~hull, for a book that explains the tools of machine learning for managers who want to work productively with data scientists.

is documented, credit decisions could be automated. This would be an example of digitization: an application of the third industrial revolution. If there was no documentation of the rules used by loan officers to make decisions, a data scientist could collect data on the features of applicants and the decisions made by loan officers so that the relationship between the decisions and the features could be learned. This would be an application of the fourth industrial revolution.

However, a data scientist could go one step further in applying the tools of the fourth industrial revolution. He or she could collect data on the features of applicants and how loans worked out (i.e., whether they were fully repaid or defaulted). Many financial institutions use machine learning in this way for routine lending decisions. It can lead to decision criteria that improve on those used by loan officers.

It is important to distinguish machine learning from traditional statistics. The difference is illustrated in Figure 29.1. In traditional statistics the first stage is to develop a hypothesis. This should be done without looking at data. Once the hypothesis has been developed and refined, data is collected to test it. Machine learning has a quite different methodology. There is no hypothesis. The first stage is to collect the data. Alternative models are fitted to the data and one is chosen.

Three types of machine learning models can be distinguished:

- Unsupervised learning models
- Supervised learning models
- Reinforcement learning models

Unsupervised learning is concerned with learning patterns in data. It is a popular tool in marketing. Consider for example a bank. It has access to many features describing its customers: average account balance, number of transactions per month, other bank services used, and so on. This could enable it to determine a number of different clusters where the customers in each cluster behave similarly. It might then be able to identify which of the bank's products (e.g., car loans, mortgages, wealth management services, etc.) each cluster is likely to be interested in and target its advertising more effectively. Two common unsupervised learning tools are the k-mean algorithm and principal components analysis

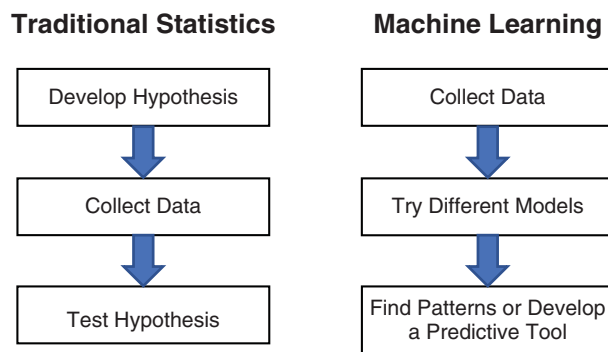


Figure 29.1 Methodologies of Statistics and Machine Learning

Supervised learning is concerned with using data to make predictions. There are two types of prediction: prediction of a value and classification. The prediction of the value of houses in a certain city is an example of value prediction. The features of houses could be square feet of living space, number of bedrooms, lot size, location, etc. Some labeled data (i.e., data where the value of the house and its features are known) must be available. The algorithm learns how the value depends on the features so that the value of other houses can be predicted.

An example of classification is the loan application example mentioned earlier. When we are attempting to model the decisions of the loan officer, the features of the applicant are used to classify the loan as “accepted” or “rejected.” When we are attempting to improve on those decisions, the features of the applicant are used to classify the loan as “fully paid back” or “defaulted.”

There are many tools for supervised learning. Linear regression (for value prediction) and logistic regression (for classification) are well known statistical techniques that are often used in machine learning. Other tools are k-nearest neighbors, decision trees, support vector machines, and neural networks.

One problem in machine learning is overfitting. If a data scientist searches widely enough, he or she may find a model that fits available data really well. But the model may not generalize to new data. For this reason, it is good practice to divide available data into three sets: a training data set, a validation data set, and a test data set. The training data set, which might be 60% of the total, is used to determine the parameters of alternative models. The validation data set (perhaps 20% of the total) is used as an out-of-sample test to see how well the models generalize to new data. The model chosen should be reasonably accurate and generalize well. The test data set (which might also be 20% of the total) is held back to provide an estimate of model accuracy.

Reinforcement learning (RL) is concerned with situations where a series of decisions have to be made in a changing environment. One application is to the hedging of derivatives.² A trader might implement a hedge today, rebalance the hedge tomorrow in the light of changing market conditions, rebalance again the next day, and so on. Reinforcement hedging is a trial-and-error tool for determining the best strategy. It underlies the algorithms that have been developed to play Go, chess, and other two-person games better than the best human players. The RL algorithm is given the rules of the game (and nothing else). It plays against itself millions of times and learns the best strategy to follow in different situations.

29.1.2 Blockchain

Ledgers recording the ownership of assets can involve the possibility of fraud, lack of trust by users, and computer hacking. Many readers will have heard stories about title fraud

²See, for example, J. Cao, J. Chen, S. Farghadani, J. Hull, Z. Poulos, Z. Wang, and J. Yuan, “Gamma and Vega Hedging Using Deep Distributional Reinforcement Learning.” Working Paper, 2022. ssrn 4106814.

where someone who is not the owner of a house manages to claim ownership and obtain a mortgage. Sometimes ledgers recording the same items are kept by different people from different organizations. There may then be inconsistencies in the records, and it can be expensive and time-consuming to sort these out.

Blockchain is a distributed ledger where many people can add entries and a community of users control how it is updated. Problems involving inconsistent records, fraud, or hacking are either impossible or extremely unlikely. Identical records are on many different computers and can be inspected at any given time. A blockchain was first developed in 2009 in connection with bitcoin, but many other applications of the technology are now being explored. The ledger defining who owns what is a long list of blocks. When a new block of transactions is created (every few minutes in the case of bitcoin), a new block is added to the chain. A constantly updated list of blocks is given to everyone who participates.

An important aspect of blockchain is what is known as hashing. A hash is a string of 64 characters. Any text can be converted into a hash, but the process cannot be reversed so that the hash is converted back into text. Even a small change in the text will totally change the hash. To illustrate this, we can go to the SHA 256 hash calculator at:

<http://www.xorbin.com/tools/sha256-hash-calculator>

Hashing the title of this book, *Risk Management and Financial Institutions* (input without italics and without hitting *Enter*), gives

fd3dce1db8e4836afde19b624bc3eee9d29c883d04f42039754b4bd440104e26

Adding the edition number so that *Risk Management and Financial Institutions 6* is input (again without italics), produces a totally different hash:

9361313d924fcc4eea0323f2292e37f2b4809d3b1296f034becb8dd0367e9a8

It is theoretically possible for two sets of records to give the same hash, but in practice this is to all intents and purposes impossible. The hash we have just looked at is formed from the numerals 0 to 9 and the first six letters of the alphabet. This means that there are $16^{64} - 1$ different hashes.³

The records in each block of a blockchain are sealed and made tamper proof with a hash. One of the inputs to the hash is the hash of the immediately preceding block. This means that any attempt to tamper with a block will be immediately spotted because it will affect that block and all subsequent blocks.

In order for the blockchain to be free from human intervention, participants must be given an incentive to check new blocks and hash them. Bitcoin does this by inviting participants, referred to as *miners*, to search for something called a *nonce*, which when added to the block will lead to a hash that starts with several zeros. The participant who finds the nonce first is rewarded with a predetermined number of bitcoins. The new block is then hashed and distributed to the computers of all participants.

³This is approximately equal to the number of atoms in the known observable universe. (In our count, we exclude a hash consisting of 64 zeros.)

There are two sorts of blockchains: permissioned and permissionless. Permissionless blockchains such as bitcoin have no restrictions on participation and there is no need for participants to know or trust each other. Permissioned blockchains are privately shared systems, and the participants may know one another and may (at least to some extent) trust one another. Permissioned blockchains are a way financial institutions can share data to speed the settlement of transactions.

Distributed ledger technologies (DLTs) such as blockchain are now receiving a great deal of attention by financial institutions and the big four accounting firms, as well as many other large businesses.

Whether blockchain and other DLT ideas live up to the hype remains to be seen. IBM has for several years been working with a number of banks to use blockchain technology to make it easier for small and medium-size businesses to trade with one another. Blockchain can make it easier to complete electronic transfers, both domestically and internationally. Other potential applications in finance are to the trading of securities, settlement, clearing, and other back-office functions. DLTs have the potential to speed up these processes. Indeed, almost real-time settlement of trades becomes a possibility. These are examples of disintermediation, which was mentioned earlier. The innovations are both a threat and an opportunity. They will reduce bank fees but may open up opportunities for new services that can be offered to consumers and businesses. Operational risks should be reduced because the possibility of fraud or cyber-attacks should be lower. This may in turn lead to a reduction in regulatory capital.

29.2 Payment Systems

Technology has already had a huge effect on the way payments are made. As a society we have moved from cash and checks to credit and debit cards to the use of mobile wallets. Developing countries have in some respects progressed further in this direction than developed countries, in part because traditional payment systems were not as well established. Many FinTech start-ups are offering new services. Some, such as PayPal, Apple Pay, and Google Wallet in the United States, and Alipay (part of Alibaba) and Tencent in China, are large, well-established companies.

The key attributes of a payments method are speed, convenience, security, simplicity, and cost. Many services such as PayPal transfer funds almost immediately. In September 2021, PayPal introduced an app offering a combination of financial tools including direct deposit, bill paying, a digital wallet, and the use of cryptocurrencies. Storing credit-card-like information in an iPhone or similar device adds to the convenience for many consumers. Indeed, given the dominant position of Apple in the smartphone market, it has been natural for Apple to expand into payments and other banking services. One can imagine that wearables such as watches or bracelets, or even implants, may be used in the same way as smartphones to add to the convenience of the way payments are made.

Security is a big issue for all forms of payment. Tens of billions of dollars are lost each year from credit card fraud. Embedding chips as well as magnetic strips in cards helps but does not eliminate this problem. We are likely to see big changes in the way fraud is avoided in the future. Digital wallets are considered to be more secure than credit cards. Many payment providers, including banks, are investigating the use of biometric authorization. Amazon One, with 60 stores across the United States, now allows a purchase to be completed by waving the palm of one's hand over a reader. Retinal scanning, facial recognition, voice authentication, and even heartbeat monitoring are all being considered. The costs of fraud are usually borne by the payments system provider, and they are of course passed on to merchants in the form of the fee charged. The merchants in turn pass the fee on to consumers. Everyone therefore has an interest in reducing fraud, and the approaches we have just discussed for making more secure payments should be welcomed.

If biometric authorization seems far-fetched, bear in mind that India has issued biometric IDs (involving fingerprints and retinal scans) to over one billion people and is laying the groundwork for a cashless society. Making more people part of the financial ecosystem is an important objective for development of the country. The biometric IDs also have the advantage that some government benefits can be distributed with less involvement from intermediaries. Of course, some would argue that the provision of biometric information is an unacceptable violation of a person's privacy, and this may slow down its acceptance in developed countries.

Some payment systems allow users to borrow money. The interest rates charged by credit card companies are very high, but it should be remembered that users do get free credit for the period of time between a purchase and the next monthly due date. (PayPal competes with this by offering 14 days of free credit.) It is likely that more convenient credit facilities, tailored to the needs of users, will be developed. Alibaba through its subsidiaries such as Alipay and Mybank is already offering many of the same services as banks.

What other services can be offered to make payment systems attractive? Many individuals remit funds on a regular basis to family members in another country. The foreign exchange services associated with those transactions are likely to get more convenient and more competitive. For businesses, easy-to-use foreign exchange hedging services that compete with those offered by banks are likely to be developed. FinTechs may carry out sophisticated analyses of sales to help a company understand its customers better or provide accounting services.

One aspect of the digitization of payments is that it becomes much easier to collect data on a person's spending habits. This could be useful to banks when making credit decisions. (Knowing how a potential borrower spends money can be almost as important as knowing how much he or she earns.) Of course it presents huge privacy issues. Sometimes payment systems sell anonymized data. Often the data are analyzed using machine learning to better understand consumer preferences.

What is the future of payments? Credit cards are likely to be increasingly replaced by mobile payments. The innovations we have mentioned will continue. Cash will probably continue to exist for some time, but the percentage of people using it will decline. It is worth noting that Sweden has led the world in moving toward becoming a cashless society, but political parties in that country are asking for regulations to require the continuation of cash payments. They are concerned about the impact of a cashless society on the elderly, lower income groups, and people in rural areas with limited or no internet service.

29.2.1 Cryptocurrencies

The use of bitcoin as a means of payment has grown rapidly since it was first introduced in 2009. Figure 29.2 shows the value of bitcoin in U.S. dollars between 2015 and June 2022. The price has been very volatile. It rose to over \$60,000 in April 2021, fell to almost \$30,000 in July 2021, rose to over \$60,000 in October 2021 before falling to less than \$20,000 in June 2022.

The advantage of using a cryptocurrency such as bitcoin as a means of payment is that it is outside the control of any one government and therefore its value cannot be debased by a government's money supply policies.⁴ This can be particularly attractive for individuals living in developing countries such as Venezuela or Zimbabwe where the value of the domestic currency is highly volatile. Payments are secure, private, and low cost. An individual can remain anonymous while using a cryptocurrency.

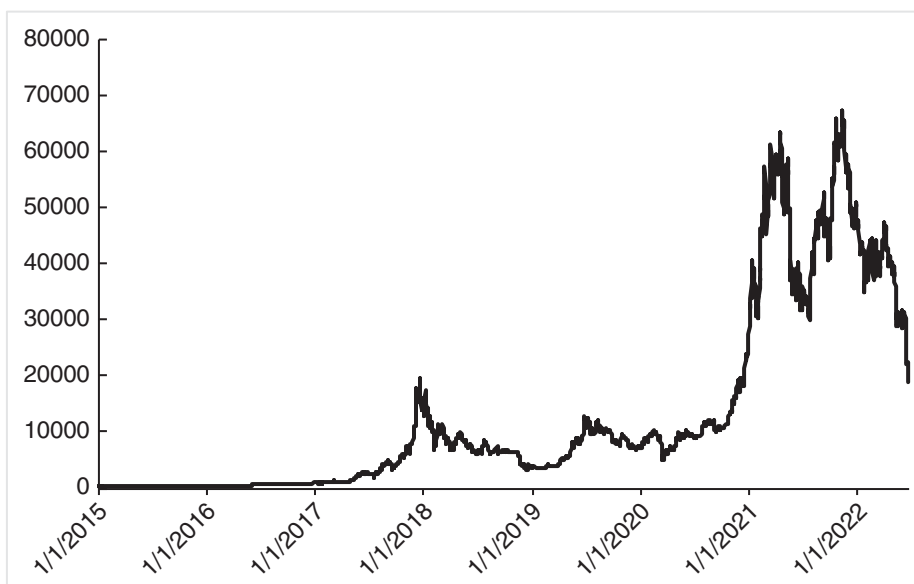


Figure 29.2 Value of Bitcoin in U.S. Dollars, January 2015 to June 2022

⁴The increase in the supply of bitcoin is determined by the amounts paid to miners (participants) when they verify transactions, as described in Section 29.1.2.

BUSINESS SNAPSHOT 29.1**Mt. Gox**

Bitcoin is a secure means of payment, but some bitcoin exchanges, which allow individuals to trade bitcoins for conventional (fiat) money, have had problems. There have been a number of reports of exchanges being hacked. The most high-profile one concerns Mt. Gox, which is based in the Shibuya district of Tokyo. In February 2014, it announced that bitcoins worth more than \$450 million were missing and likely stolen. The reason for the loss, which happened over a number of years, is unclear. It appears that bad management and poorly controlled computer code allowed hackers to steal bitcoins. Some have blamed the CEO for embezzling the funds, but this has been denied. Whatever the reason, Mt. Gox filed for bankruptcy in April 2014.

How can you avoid these types of losses? Bitcoins should not be stored on an exchange. They should be transferred to a trustworthy online exchange only when you want to trade them.

The disadvantage of a cryptocurrency is that its value is more volatile than other fiat (traditional) currencies.⁵ It pays no interest. It is often associated (fairly or unfairly) with illegal activities such as drug dealing. Jamie Dimon, CEO of JPMorgan, is on record as saying “Bitcoin is a fraud that will ultimately blow up.” China has introduced laws banning the use of bitcoin and the activities of bitcoin exchanges. Users of cryptocurrencies are reliant on the way a cryptocurrency has been set up. If a flaw were found, the currency would be worthless. The bitcoin protocol has not been successfully hacked, but bitcoin exchanges have been. (See Business Snapshot 29.1.)

Bitcoin (BTC) was the first cryptocurrency, but many others have been developed. The largest by market capitalization in 2022 were: Ethereum (ETH), Tether (USDT), USD Coin (USDC), Binance Coin (BNB), Binance USD (BUSD), Cardano (ADA), XRP (XRP), Solana (SOL), Dogecoin (DOGE), Polkadot (DOT), and Dai (DAI). Some have procedures for making transactions secure that are quite different from those used by bitcoin. Some cryptocurrencies such as ethereum facilitate *smart contracts*. These are contracts that execute automatically when specified conditions are met, thereby eliminating the need for intermediaries.

A stablecoin is a particular type of cryptocurrency whose value is tied to an outside asset such as the U.S. dollar or gold. A number of the cryptocurrencies mentioned earlier

⁵The volatility of the U.S. dollar value of bitcoin calculated from the data in Figure 29.2 is about 62%. The volatility of a fiat currency, such as the euro, British pound, or Japanese yen when expressed in terms of the U.S. dollar, is typically less than 20% per year.

(e.g., Tether and USD Coin) are stablecoins. They maintain their value by holding the outside asset. When a stable coin is bought, an equivalent amount of the outside asset is bought. When it is sold, an equivalent amount of the outside asset is sold. Stablecoins are designed to have the advantages of cryptocurrencies, such as bitcoin, without the volatility.

Some cryptocurrencies have tried to use algorithms to match the value of a stablecoin to the outside asset. They do this by increasing or decreasing the supply of the stablecoin. In 2022, the stablecoin TerraUSD and its sister token, Luna, which used algorithms, failed spectacularly, leading to \$60 billion of losses. Algorithmic strategies for creating stablecoins may never become popular with investors again.

29.2.2 Central Bank Digital Currencies

Nearly all major countries of the world are exploring the possibility of offering their own digital currencies. These are known as central bank digital currencies (CBDCs). Instead of issuing coins and notes that are backed by the credit of the government, the central bank issues electronic money, which is similarly backed by the credit of the government. By 2022, a number of small countries, such as Jamaica and Nigeria, had issued CBDCs. It is expected that India, China, and Russia will become the first large countries to do so. Other countries, such as the United States and the U.K., are carrying out research on CBDCs, but moving more slowly. One advantage of CBDCs is that they can provide easier and safer access to money for the unbanked. They can lead to more efficient domestic payments and lower transaction costs. They may also make it easier for the central bank to control the money supply.

One can imagine a country where all citizens have a digital currency account at the central bank and private companies compete to provide payment services. The central bank can increase the money supply by adding to the holdings of all citizens in an equitable way.

Some of the implications of all this are of course uncertain. Some international standard setting is likely to be necessary so that a digital currency of one country can be converted to that of another. Replacing banknotes with digital currencies would make it much easier for a resident of one country to transact in the digital currency of another country. This could lead to a new form of competition between the currencies of different countries and could actually end up making it more difficult for any one country to manage its economy by controlling the money supply.

29.3 Open Banking

Open banking is banking practice that gives third-party financial service providers (often relatively small FinTechs) access to a bank's customer data through the use of application programming interfaces (APIs). Customers must provide consent to their data being transferred in this way. The advantage to them is that they have easy access to services offered by different organizations.

The U.K. has been a leader in open banking. At the end of 2021 there were 4.5 million regular users of open banking in the U.K. The Payments Services Directive 2 of the European Union (PSD 2) required banks to make open banking services available to their customers. Other countries, such as the United States, are expected to introduce open banking during the first half of the 2020s. The key advantage of open banking is that it encourages financial innovation and gives consumers access to a wider range of financial services. However, some banks may feel threatened by it as it increases competition and may lead to some lost business.

29.4 Lending

As all students of banking know, there are a number of Cs associated with lending: character, capacity, capital, collateral, and so on. Machine learning (mentioned earlier) has been quite successful in automating lending decisions. In principle, a machine learning program can be more objective and exhibit less bias than a human being.⁶ We now consider two innovations in lending, neither of which have proved a threat to the activities of banks in this area.

29.4.1 P2P Lending

A bank's net interest margin is determined by the difference between the interest it pays on liabilities and the interest it earns on assets. If it obtains deposits on which it is paying 2% per annum and then lends the funds out at 5% per annum, its net interest margin is 3%. Some of this (perhaps as much as 1%) is to compensate for expected loan losses, but if the depositor could lend money directly to the borrower at, say, 4%, both sides could potentially be better off. Peer-to-peer (P2P) lending involves disintermediation followed by reintermediation. The new intermediaries are set up to provide services such as:

1. Providing online platforms for borrowers and lenders to contact each other.
2. Verifying the borrower's identity, bank account, employment, income, and so on.
3. Assessing the borrower's credit risk and, if the borrower is approved, determining the appropriate interest rate.
4. Processing cash flows between borrowers and lenders.
5. Attempting to collect payments from borrowers who are in default.

Many borrowers who currently use P2P platforms have already been refused by banks, so the interest rates can be quite high compared with conventional loans (but lower than the rates on credit card balances and other sources of credit for moderate- to high-risk borrowers). Examples of PTP lenders are Kiva, Prosper, Funding Circle, and Peerform. The P2P lending platform typically assigns a credit rating to a borrower in much the same way that a bank does.

⁶But it has been shown that machine learning is not completely free from bias. The data that are used by the machine learning algorithm may incorporate biases. Alternatively, the machine learning programmer may consciously or unconsciously incorporate biases into the way algorithms are used and trade-offs are made.

It clearly makes sense for lenders to be diversified. Typically, they can either choose a portfolio of loans themselves or ask the P2P lending platform to construct a portfolio that is consistent with their risk preferences. The risk of borrowers is often idiosyncratic so that correlations between the performances of different loans are low, making diversification easier.

Some P2P platforms allow lenders to sell a loan before it has been fully repaid, but this depends on the willingness of another client of the P2P lending platform to take over the loan. There is likely to be a natural assumption that if a lender wants to sell a loan there is something wrong with it.⁷ However, it may just be that the lender has run into liquidity problems so that he or she needs cash. Creating a better secondary market for loans is a way the market can evolve. An idea that can be used for increasing liquidity is to create tranches from portfolios of loans as described in Chapter 7 and sell securities created from them to lenders.

29.4.2 Crowdfunding

Crowdfunding is another activity that has the potential to be a disruptive force in finance. It involves obtaining funds for a project from a large number of different investors. Some crowdfunding projects are donation-based with the purpose of funding projects that donors identify with and want to see funded. Other crowdfunding platforms are equity-based, designed to provide investors with an ownership interest in a company that might become successful. Sometimes donors receive a service or physical product in exchange for a contribution.

Crowdfunding has opened up equity investing in private companies to a much wider range of investors than venture capitalists and angel investors. The valuation of the company is a critical part of equity crowdfunding. If a company is valued at \$500,000, then shares representing 1% of the company should require an investment of \$5,000; if it is valued at \$1 million, the required investment is \$10,000. Sometimes the company determines the value and sometimes investors do so. In the latter case, some platforms use experienced professionals to carry out due diligence and ensure that an investment is properly priced. Ideally, the professional is an angel investor who has some skin in the game. Of course, he or she will typically require to be compensated in some way, and it should be recognized that, for even an experienced professional, valuations are highly subjective.⁸

The growth of equity crowdfunding has been slowed by regulation. Most countries have strict rules on the issuance of debt and equity securities. In the United States, the Jumpstart Our Business Startups (JOBS) Act of 2012 provides exemptions from the rules for certain crowdfunding activities. Under the JOBS Act, the maximum amount of funding a company can raise in a year is limited to about \$1 million, and the maximum funds an investor can provide is subject to a limit based on the investor's income.

⁷Economists refer to this as the “market for lemons” problem following a 1970 paper by George Akerlof with that name.

⁸As entertaining television shows such as *Shark Tank* illustrate, experts do not always agree when valuing early-stage businesses.

Initial coin offerings (ICOs) are a form of crowdfunding involving cryptocurrencies. Typically a company issues to investors its own cryptocurrency in exchange for an established one such as bitcoin or ethereum. The investors have no claim on the company, but if the company does well, they expect the cryptocurrency it has issued to increase in value. This is because use of the new cryptocurrency is integral to the operations of the company. ICOs have attracted regulatory attention. Many ICOs have been scams. But they have launched a few successful companies. China has banned them, and other countries are considering whether they should be regulated in the same way as IPOs.

29.5 Wealth Management

Wealth management has traditionally been very profitable for banks. Fees are often in the 1% to 1.5% range of the amount invested per year and can be much more when hidden fees associated with mutual fund investments and trading costs are taken into account. Once a client's risk appetite has been assessed, wealth management involves finding appropriate investments for the client. Often similar investments are recommended to all clients with similar risk appetites, easing the job of being a wealth manager. It is not surprising that wealth management is an industry that is in the process of being disrupted. As mentioned in Chapter 4, John Bogle took the first step toward reducing the costs of investing with the first index fund in 1975. Index funds have since become very popular and charge fees as low as 0.15% with no human intervention in the form of a wealth manager being necessary.

Robo-advisers first appeared in about 2010. In most countries they must register with the authorities and are subject to regulation. They provide digital platforms where investors express their risk preferences. A portfolio, usually involving ETFs tracking an equity index and safe investments, is then chosen, and the portfolio is automatically rebalanced as necessary. There is very little human intervention. Fees are lower than those charged by traditional wealth managers, typically often only 0.25% of the amount invested per year. Well-known companies include Wealthfront and Betterment. Some banks and other wealth managers are now responding to the competition by offering their own automated wealth management services. Indeed, those that fail to do this are unlikely to survive. Providers of index mutual funds, such as Vanguard, are also active in this space.

Robo-advisers are making investment advice available to a much wider range of individuals than previously. Often investors can start with as little as \$500 or \$1,000 whereas a traditional wealth manager might require a minimum investment of \$50,000. Robo-advisers make it easy for clients to add to the funds under management on a regular basis. Arguably they serve an important role in society by encouraging people to save when they might not otherwise do so. Robo-advising tended to attract young investors with small amounts to invest in its early days, but a much wider range of investors, including those classified as “high net worth” and HENRYs (high earners not rich yet) are now using the services.

Up to now the main innovation underlying robo-advising has been the delivery of services in a cheaper, novel way that many investors find appealing. The investment strategies underlying the advice given are usually similar to those that have been used by the investment industry for many years (see Chapter 1). Tax-related strategies (such as tax loss harvesting) are often incorporated into the advice that is given.⁹

There is plenty of scope for the strategies to become more sophisticated. Investments can be better diversified internationally and across sectors. They can be better targeted to the goals of the investor, taking into account the investor's age, retirement plans, plans to pay for the higher education of children, and so on.

In 1992, Fischer Black and Robert Litterman at Goldman Sachs published a widely used way of incorporating the views of investors in the selection of a portfolio.¹⁰ Robo-advisers may find ways of expanding the range of alternatives offered to investors using this technology. Alternative sets of views with rationales could be presented, with investors being invited to choose between them. It might even be possible to let the views of the investor be a less structured direct input to the determination of the portfolio.

In Chapter 28, we mentioned that a number of cognitive biases influence decision making. Investors are subject to these biases. For example, they are reluctant to sell losers, they chase trends, and they get disillusioned and exit equity markets when they should stay for the long term. It is often the ability to avoid these biases that distinguishes a professional investor from an amateur. Robo-advisers can try to stop investors falling victim to the biases by developing innovative ways of explaining them. Finally, robo-advising can be combined with other financial innovations so that a percentage of a client's funds is allocated to P2P lending or equity crowdfunding.

Robo-advising has already become an important part of the financial landscape and is likely to become more widely used as the millennial generation accumulates wealth. For this generation, it is much cooler to invest with an iPhone than make a trip to the bank. However, it is perhaps worth sounding one note of caution. Equity markets performed really well in the years following the start of robo-advising in 2010. The appeal of robo-advising may decline when there is a downturn and the clients of robo-advisers, who never invested before, complain about losing money. It is hoped that the advisers will be able to educate investors on the importance of staying focused on the long term.

29.6 InsurTech

The application of technology in the insurance industry is referred to as *InsurTech*. In some respects insurance companies are less likely to be disrupted by technology innovations than banks are. This is because the nature of insurance is such that consumers

⁹Tax loss harvesting involves selling stocks with capital losses at the year end to reduce the capital gains tax payable.

¹⁰See F. Black and R. Litterman, "Global Portfolio Optimization," *Financial Analysts Journal* 48, no. 5 (1992): 28–43.

require a conservatively run, stable company. For example, it does not make sense to buy house insurance if there is a non-negligible probability that the insurance company will not be around in the event of the house being destroyed by fire. Regulators monitor the financial condition of insurance companies carefully, and it is very difficult for a new company to offer insurance unless it has a large amount of capital to cover claims. However, some applications of technology in insurance are straightforward. For example, it is now easy to compare the quotes of many different insurance companies online.

Often customers do not deal directly with insurance companies. They go through brokers or agents. These intermediaries are sometimes tied to just one insurance company or to a small number of insurance companies, rather than having a free hand to find the best deal for clients. A natural development is disintermediation where insurance companies create websites for dealing directly with customers. As the market develops, brokers and agents are likely to either disappear or become digitized providers of price comparisons.

One important potential InsurTech application is the collection of data by insurance companies so that risks are assessed more accurately. Auto insurance is a good example. By placing a “black box” in a car, an insurance company can use telematics (the branch of information technology that deals with the long-distance transmission of computerized information) to accumulate data on a client’s driving. The black box can record driving speeds, distances traveled, the time of day or night that the car is being driven, driving style (e.g., frequency of braking and the cornering speed), and the type of road usually used. This enables the insurance company to build up an accurate profile of the driver. If the driver is less risky than average, the insurance premium charged will go down. Similarly, if the driver is more risky than average, it will go up.

The black box can be used for driver education. The driver may be able to access a secure website to find out how well he or she is performing. This can suggest changes to driving style and provide tips on bringing down the cost of insurance. In principle, this could improve the average ability of drivers and make roads safer. Of course not everyone can expect a reduced premium as a result of a black box, and some drivers may regard them as an invasion of privacy. One can speculate that as black boxes become more widely accepted, only good drivers will use them and the premiums for non-users will increase.

To the insurance company, the key advantage of a black box is that it is able to collect data allowing risks to be assessed more accurately. In an ideal world, an insurance premium charged should reflect miles driven and the risks incurred by the driver per mile.¹¹ Another potential advantage of black boxes to the insurance company is that they may be able to pinpoint blame in the event of an accident and make claims simpler.

¹¹Metromile, a San Francisco company, is an example of an insurance company moving in this direction. It uses a gadget to monitor mileage driven and sends a monthly invoice for insurance to policy holders to reflect the records of miles driven. The policies are actually written by a large well-capitalized and highly regulated insurance company, National General Insurance.

A general point is that insurance companies now have opportunities to collect more data than ever before to assess the risks they are taking. These data can potentially allow them to improve estimates of an individual's life expectancy, the probability of a particular house burning down, or the chance of a hurricane in a certain part of the world. Machine learning provides the tools to analyze the data. The insurance industry is by nature conservative (perhaps rightly so), and it is difficult to imagine InsurTech radically changing the structure of the industry. (InsurTech companies usually have to partner with established insurers.) But the companies that prosper are likely to be those that take advantage of the opportunities created by new data sources.

29.7 Regulation and Compliance

There are two aspects of the impact of technology on regulation. One is how regulators should handle companies that use technology in innovative ways and have the potential to be a disruptive force in finance. The other, known as *RegTech*, is concerned with how regulation itself can use technology. We will consider both in this section.

29.7.1 Regulation of Financial Innovators

As we have seen in this book, financial institutions are highly regulated. This is a barrier to entry for small FinTech start-ups. For example, P2P lenders in the United States must enter into an arrangement with a bank to comply with regulations. Robo-advisers must satisfy know-your-client (KYC) rules, even though there is limited or no human interaction in their services, and register with the authorities. If a FinTech ignores regulations, it is likely to be closed down by supervisory authorities, at least temporarily, and suffer reputational damage.

Regulators do not want to stifle financial innovation. Indeed, as we have pointed out, financial innovation can bring many benefits to society by developing innovative ways of saving, investing, and borrowing so that more people are encouraged to be part of the financial ecosystem. But regulators must tread a fine line. They do not want *shadow banks* (i.e., non-banks that are offering similar services to banks) to subject clients to unreasonable risks.

Some regulatory authorities have initiated “sandbox” programs.¹² A sandbox is a safe space in which businesses can test innovative products, services, business models, and delivery mechanisms without immediately incurring all the normal regulatory consequences of pilot activities.

Regulating FinTech companies is likely to be a challenge for regulators in the future. The main asset of a FinTech company is often its computer code, and it is not at all easy for supervisors to evaluate this in the same way that they would evaluate a bank's

¹²One of the first was in the U.K. See Financial Conduct Authority, “Regulatory Sandbox,” November 2015.

systems. High-frequency traders (HFTs) illustrate the problem. Their trading activities are determined entirely by their computer code. On May 6, 2010, largely as a result of the activities of HFTs, a “flash crash” occurred in the United States where the Dow Jones Industrial Average declined by about 9% in a matter of minutes before partially recovering. On August 1, 2012, a software error in the program used by an HFT caused a major disruption in the prices of 148 stocks listed on the New York Stock Exchange. The Securities and Exchange Commission has taken some steps to avoid similar problems in the future, but it is difficult for it to guard against all the ways in which software can (intentionally or unintentionally) interfere adversely with the functioning of financial markets.

Another issue for regulators is that the jurisdiction may be difficult to determine for some FinTech companies. This is particularly likely to be the case if all of its transactions are carried out in digital currencies.

29.7.2 RegTech

RegTech is the application of technology to regulation and compliance. Some of the applications are straightforward. Regulatory reporting has traditionally involved thousands of pages of documents being provided to regulators at the end of each reporting period (e.g., every quarter). Moving this online and providing data in real time have the potential to simplify the job of both the regulator and the regulated financial institution, while making regulation more timely and efficient.

Fines for failing to comply with regulations can be steep.¹³ For example, in Chapter 20 we mentioned that a French bank, BNP Paribas, was fined an amount equal to its profits for a whole year for violating U.S. sanctions. No doubt this happened because the activities of a small number of employees were not being monitored closely by the bank’s compliance group. Technology can make it easier for a bank to ensure that it complies with regulations and the laws of the different countries it operates in. For example, technology can make it possible for it to screen both new customers and new transactions in real time to identify problems. Compliance violations involving issues such as drug money laundering, terrorist financing, and sanctions can be effectively eliminated if a real-time RegTech application, approved by regulators and reflecting the latest rules, is used.

Preventing staff from engaging in questionable behavior is important to financial institutions, as the behavior can prove to be very costly in terms of regulatory fines, legal costs, and reputation. Some innovations allow companies to keep track of their employees’ conversations across a number of communications platforms. For example, Digital Reasoning, a company headquartered in Nashville, with offices in London, New York, and Washington, has created surveillance software used by some of the largest banks and asset management companies. It can monitor employee behavior by analyzing

¹³Since the Global Financial Crisis, banks throughout the world have paid hundreds of billions of dollars in fines. Technology to improve a bank’s compliance can therefore have a huge payoff.

millions of e-mails, chat logs, and phone conversations to detect suspicious or unusual activity using machine learning. It can handle six different English dialects. If machine learning detects an employee whose behavior is quite different from the usual pattern, further investigation might be warranted. In an asset management firm, the behavior might prove to be indicative of insider trading, which, if allowed to continue, could lead to a huge fine. In a bank, it might prove to be indicative of rogue trading or a failure to treat subordinates with respect.

As mentioned in Chapter 20, cyber risk is the biggest operational risk for banks. It is also becoming a regulatory issue. For example, the General Data Protection Regulation in Europe leads to financial institutions being charged hefty fines if they lose data meant to remain confidential. Some RegTech applications are helping banks keep their data secure and comply with regulations like this one.

Just knowing all the regulations that exist is a major problem for financial institutions, as the regulatory environment it operates in throughout the world changes almost daily. IBM's Watson Regulatory Compliance gives relevant employees access to a library of regulatory requirements in which materials can be filtered geographically, by business type, by product, and by compliance area. This application can automatically highlight the parts of a new regulation that are relevant to a particular issue and compare the regulations related to a particular activity in different parts of the world.

As explained in earlier chapters, regulation has moved from being unidimensional to being multidimensional. Regulators used to focus on only a single capital measure. They now have two capital measures (one based on risk-weighted assets and the other based on a simpler leverage calculation) and two liquidity ratios (the liquidity coverage ratio and the net stable funding ratio). Different activities undertaken by the bank have different effects on the ratios. This leads to the potential for technology firms to come up with ways of optimizing bank strategies or providing convenient scenario analysis tools.

29.8 How Should Financial Institutions Respond?

FinTech innovations pose a strategic risk for large banks and other financial institutions. Many of the services they offer will be disrupted. How should they respond? Clearly they should embrace technological change, rather than hoping it is a fad that will disappear. But they should also carefully evaluate how consumer behavior is being affected by technological change and take steps to change their business model accordingly.

Business Snapshot 29.2 describes Eastman Kodak, a company that did not survive technology changes well. A key point is that Kodak was not unaware of the changes taking place in its industry. Indeed, the first digital camera was created in 1975 by an engineer working for Kodak, and Kodak invested billions in the new technology. So where did Kodak go wrong? It understood the new technology but did not appreciate the way it was changing consumer behavior until too late. In the course of a short period of time, a large segment of the population (not just young people) became comfortable sharing photographs and storing them on computers. They no longer felt the need to print photographs.

BUSINESS SNAPSHOT 29.2**Eastman Kodak and the Digitization of Photography**

In the 1980s and 1990s, the Eastman Kodak company was a successful business selling cameras and film for cameras. It had almost single-handedly transformed photography in the United States from something that was the exclusive preserve of professional photographers to something that almost everyone did. It had successfully navigated technological changes in its industry. It moved from dry-plate to black-and-white film and then sensibly invested in color film even when it was demonstrably inferior to the black-and-white film from which it derived most of its revenue. Its market capitalization in 1997 was about \$30 billion.

Steve Sasson, an engineer at Kodak, invented the first digital camera in 1975. According to Sasson, the company's response was "That's cute—but don't tell anyone about it." In fact, Kodak did not ignore the trend to digital cameras. It invested heavily and in 1995 it brought its first digital camera to market, the DC40. In 2001, it launched the EasyShare line of point-and-shoot cameras. Kodak had a first-rate research capability. It had the ability to develop state-of-the-art cameras that could enhance colors, automatically adjust lighting, organize portfolios, share photos, detect smiles to get the best shots, and so on.

Unfortunately, Kodak exhibited some reluctance to continually improve products and innovate, a mistake in an industry where technology changes fast. Cannibalizing existing products in the name of progress is painful, but necessary (even when the new products are less profitable than the old ones). How well Kodak would have survived the smartphone revolution if it had adopted a different strategy is debatable. All we can say is that some of its competitors such as Fuji have fared better. In 2022, Fuji's market capitalization was more than \$18 billion. Kodak filed for bankruptcy at the end of 2012, and its successor company had a market capitalization of about \$400 million in 2022.

Kodak coined the term "Kodak moment," which it used extensively in its advertising to convince people that they should have a Kodak camera loaded with Kodak film ready to use at all times. It is easy to be wise after the event, but strategy professionals would argue that the company could have extrapolated from its sales pitch to recognize the business it was in. It was in the imaging or moment-sharing business rather than the film business. We will never know how well it would have survived if it had more enthusiastically embraced electronic sharing and the storing of digital images when the market was in its early stages. (As indicated in Business Snapshot 29.2, Fuji, which was a similar company to Kodak but based in Japan, fared better than Kodak.) It was its implicit

belief that demand for hard-copy photographs would continue that ultimately doomed the company.

The disruption of large financial institutions does not seem to be happening as quickly as the disruption of Kodak. They have a number of competitive advantages. They are well capitalized (although the same is true of Apple, Google, and Alibaba). They understand how to deal with the highly regulated environment they operate in (something many FinTech start-ups find difficult). They have a huge customer base and are still trusted by many customers (although it is important to recognize that the 2007–2008 financial crisis and events since then such as the Wells Fargo debacle described in Business Snapshot 24.1 have served to erode this trust). One can speculate that financial institutions are not as vulnerable as Kodak in that many people are less inclined to experiment with the way their money is handled than with the way they do photography. Also, many start-ups need established financial institutions to offer their products.

However, there are some warning signs that banks should respond to. A Millennial Disruption Index survey indicated that 71% of millennials in the U.S. would rather visit the dentist than listen to what banks are saying, while 73% would rather handle their financial services needs through Google, Amazon, Apple, PayPal, or Square.¹⁴ Millennials also voted four leading U.S. banks among their least loved brands. Millennials are individuals born between 1981 and 2000 and are important future consumers of bank services. In 2022 they made up about 40% of the workforce and by 2050 they will have inherited about \$30 trillion. They may feel less radically inclined about financial services as they age, but banks should not rely on this.

Kodak was ultimately rendered largely irrelevant by the digital cameras incorporated in smartphones. Financial institutions naturally do not want to become similarly irrelevant. They have recognized the need to offer mobile apps for payments, wealth management, and a host of other services. But it is important for them to embrace technological change, not just to pay lip service to it. Technological change in the financial sector will continue at an accelerating rate. In many cases it will erode the profits banks previously relied upon (as was the case for Kodak). Being flexible enough to adjust will be a continuing challenge.

The new services developed by banks need to be convenient and designed so that young people classify them as “cool” while older people find them easy to use. There are a number of different approaches being adopted. Some financial institutions have developed new services in-house; some have bought start-ups that have already developed the services; and some have entered into partnerships with start-ups. The first alternative, although the least expensive and most appealing to many working in the financial services sector, can be quite difficult given the somewhat complacent culture that often permeates large companies. The second and third alternatives can be used as a way of disrupting the culture and accelerating change. Some banks have found it useful to create an organizationally distinct unit that has the ability to bring in outside talent when

¹⁴See www.millennialdisruptionindex.com.

necessary and can partner with start-ups. Whatever approach is adopted, strong support from the CEO and the board is necessary for success.

The financial institutions that prosper will be the ones that manage to change their culture so that they are able to provide services in a fast, flexible way that appeals to consumers. Lou Gerstner was appointed CEO at IBM in the 1990s when the company was losing money. He successfully changed the culture and transformed it into a profitable company. He is quoted as saying, “Culture isn’t just one aspect of the game. It is the game.” He successfully changed IBM from being a supplier of mainframe computers to being a business problem solver where teams of people worked well together. Banks are similar to IBM in that they are huge organizations with many bright employees. The job of the CEO is to create a culture where individuals are rewarded for embracing technology and finding innovative ways of making their services more convenient and valuable for clients.

Summary

What will the financial services sector look like in, say, 2040? Of course, the pace of technological change is so fast that accurate forecasts are impossible. However, offering a few predictions is a convenient way of summarizing some of the points made in this chapter.

Checks and credit cards will largely disappear. Mobile wallets on smartphones and wearables will become the norm. Biometrics will be used to make payments more secure. Some central banks will choose to switch from paper currencies to digital currencies, and transacting in digital currencies will become more widespread. Machine learning will be able to undertake many tasks such as credit assessment and fraud detection much better than human beings. Record keeping will be more secure and faster using blockchains and other DLTs.

Large insurance companies will continue to exist, but brokers and agents will not. Insurers will have access to more data than ever before and will use machine learning and other techniques to assess risks more precisely. There will be partnerships formed between insurance companies and technology firms to deliver insurance services in innovative ways.

Investing and borrowing will see big changes. Crowdfunding and P2P lending may become more widely used. Some of today’s large banks will find it necessary to offer these services to savers and borrowers. It will be unusual for funds to remain idle for any length of time, and the ease of borrowing and investing will make more people part of the financial ecosystem.

Some banks that fail to adjust to new technology enthusiastically will fail and be acquired by other banks. (Indeed, too-big-to-fail may no longer be considered a relevant descriptor of very large banks.) Some large technology firms will become banks and offer a full range of banking services. Most bank services will be offered more cheaply and more efficiently than before so that banks will lose some of their traditional sources of profit. The banks that survive will have to cut costs by making big reductions in the

number of branches they have and the number of people they employ. Banks will have to form partnerships with many different technology firms, to keep the services they offer up to date.

Further Reading

Chishti, S., and J. Barberis. *The FinTech Book: The Financial Technology Handbook for Investors, Entrepreneurs, and Visionaries*. Chichester, UK: John Wiley & Sons, 2016.

Christensen, C. M., and M. E. Raynor. *The Innovator's Solution*. Boston: Harvard Business Review Press, 2003.

Harvey, C. R. "Cryptofinance." Working Paper, SSRN 2438299.

IOSCO. *Research Report on Financial Technologies*, February 2017.

Sironi, P. *FinTech Innovation: From Robo-Advisors to Goal-Based Investing and Gamification*. Chichester, UK: John Wiley & Sons, 2016.

Practice Questions and Problems (Answers at End of Book)

- 29.1 Explain the terms *disintermediation* and *reintermediation*.
- 29.2 What is meant by (a) machine learning and (b) distributed ledger technologies?
- 29.3 What is meant by biometric authorization? Give some examples.
- 29.4 Why might a central bank switch from fiat money to digital money?
- 29.5 Explain how P2P lending and equity crowdfunding work.
- 29.6 "When thinking about the impact of FinTech, financial institutions should be thankful for regulation." Explain.
- 29.7 "Index funds are a great idea, but if everyone chose them there would be no price discovery." Do you agree with this argument?
- 29.8 What is meant by RegTech? Give some examples.
- 29.9 "Banks should aim to respond to change like IBM, not like Kodak." Discuss this statement.

Further Questions

- 29.10 Explain how machine learning is changing the way lending decisions are made. What are the risks in relying on machine learning algorithms for these decisions?
- 29.11 With the benefit of hindsight, we can say that Kodak was in the imaging and moment-sharing business. What business are banks in?

Chapter 30

Risk Management Mistakes to Avoid

Since the mid-1980s, there have been some spectacular losses in financial markets. This chapter uses them to review some of the key points made in this book. Some of the losses are listed in Business Snapshot 30.1.

One remarkable aspect of the list in Business Snapshot 30.1 is the number of times huge losses were caused by the activities of a single person. In 1995, Nick Leeson's trading brought a 200-year-old British bank, Barings, to its knees; in 1994, Robert Citron's trading led to Orange County, a municipality in California, losing about \$2 billion. Joseph Jett's trading for Kidder Peabody caused losses of \$350 million. John Rusnak's losses of \$700 million at Allied Irish Bank came to light in 2002. Jérôme Kerviel lost over \$7 billion for Société Générale in 2008. Kweku Adoboli lost \$2.3 billion for UBS in 2011.

The losses emphasize the importance of internal controls. Many of the losses we will consider occurred because systems were inadequate so that the risky positions being taken were simply not known. It is also important for risk managers to continually "think outside the box" about what could go wrong so that as many potential adverse events as possible are identified.

BUSINESS SNAPSHOT 30.1

Big Losses

Allied Irish Bank: This bank lost about \$700 million from the unauthorized speculative activities of one of its foreign exchange traders, John Rusnak, which lasted for a number of years. Rusnak covered up his losses by creating fictitious options trades.

Archegos Capital Management: This investment office managed the personal assets of Bill Hwang. It was highly levered and dealt with multiple banks, each of which was likely unaware of Archegos' total exposure. On March 26, 2021, after its investments had lost money, Archegos defaulted on margin calls from several banks. Credit Suisse lost about \$5.5 billion; Nomura lost about \$2.85 billion; Morgan Stanley, UBS and Mitsubishi incurred smaller losses.

Barings: This 200-year-old British bank was wiped out in 1995 by the activities of one trader, Nick Leeson, in Singapore. The trader's mandate was to arbitrage between Nikkei 225 futures quotes in Singapore and Osaka. Instead he made big bets on the future direction of the Nikkei 225 using futures and options. The total loss was close to \$1 billion.

Enron's Counterparties: Enron managed to conceal its true situation from its shareholders with some creative contracts. Several financial institutions that allegedly helped Enron do this have each had to settle shareholder lawsuits for over \$1 billion.

Hammersmith and Fulham (see Business Snapshot 20.1): This British Local Authority lost about \$600 million on sterling interest rate swaps and options in 1988. The two traders responsible for the loss knew surprisingly little about the products they were trading.

Kidder Peabody (see Business Snapshot 22.2): The activities of a single trader, Joseph Jett, led to this New York investment dealer losing \$350 million trading U.S. government securities. The loss arose because of a mistake in the way the company's computer system calculated profits.

Long-Term Capital Management (see Business Snapshot 16.1): This hedge fund lost about \$4 billion in 1998 carrying out convergence arbitrage strategies. The loss was caused by a flight to quality after Russia defaulted on its debt.

National Westminster Bank: This British bank lost about \$130 million from using an inappropriate model to value swap options in 1997.

Orange County (see Appendix B): The activities of the treasurer, Robert Citron, led to this California municipality losing about \$2 billion in 1994. The treasurer was using derivatives to speculate that interest rates would not rise.

Procter & Gamble (see Business Snapshot 5.4): The treasury department of this large U.S. company lost about \$90 million in 1994 trading highly exotic

interest rate derivatives contracts with Bankers Trust. It later sued Bankers Trust and settled out of court.

Société Générale (see Business Snapshot 5.5): Jérôme Kerviel, an equity trader in the Paris office, lost over \$7 billion speculating on movements in equity indices in January 2008. He is alleged to have concealed his exposure by creating fictitious trades. Like Leeson at Barings, his mandate was to do arbitrage trades.

Subprime Mortgage Losses (see Chapter 7): In 2007, investors lost confidence in the structured products created from U.S. subprime mortgages. This led to a “credit crunch” with losses of tens of billions of dollars by financial institutions and the worst recession since the 1930s.

UBS: In 2011, Kweku Adoboli lost \$2.3 billion taking unauthorized speculative positions in stock market indices.

30.1 Risk Limits

The first and most important lesson from the losses concerns risk limits. It is essential that all companies (financial and nonfinancial) define in a clear and unambiguous way limits to the financial risks that can be taken. They should then set up procedures for ensuring that the limits are adhered to. Ideally, overall risk limits should be set at board level. These should then be converted to limits applicable to the individuals responsible for managing particular risks. Daily reports should indicate the gain or loss that will be experienced for particular movements in market variables. These should be checked against the actual gains and losses that are experienced to ensure that the valuation procedures underlying the reports are accurate.

It is particularly important that companies monitor risks carefully when derivatives are used. This is because derivatives can be used for hedging or speculation or arbitrage. Without close monitoring, it is impossible to know whether a derivatives trader has switched from being a hedger to a speculator or switched from being an arbitrageur to being a speculator. The Barings, Société Générale, and UBS losses are classic examples of what can go wrong. In each case, the trader’s mandate was to carry out low-risk arbitrage and hedging. Unknown to their superiors, they switched from being arbitrageurs and hedgers to taking huge bets on the future direction of market variables. The systems within the banks were inadequate and did not detect what was going on.

The argument here is not that no risks should be taken. A trader in a financial institution or a fund manager should be allowed to take positions on the future direction of relevant market variables. What we are arguing is that the sizes of the positions that can be taken should be limited and the systems in place should accurately report the risks being taken.

30.1.1 A Difficult Situation

What happens if an individual exceeds risk limits and makes a profit? This is a tricky issue for senior management. It is tempting to ignore violations of risk limits when profits result. However, this is shortsighted. It leads to a culture where risk limits are not taken seriously, and it paves the way for a disaster. The classic example here is Orange County. Robert Citron's activities in 1991–1993 had been very profitable for Orange County, and the municipality had come to rely on his trading for additional funding. People chose to ignore the risks he was taking because he had produced profits. Unfortunately, the losses made in 1994 far exceeded the profits from previous years.

The penalties for exceeding risk limits should be just as great when profits result as when losses result. Otherwise, traders who make losses are liable to keep increasing their bets in the hope that eventually a profit will result and all will be forgiven.

30.1.2 Do Not Assume You Can Outguess the Market

Some traders are quite possibly better than others. But no trader gets it right all the time. A trader who correctly predicts the direction in which market variables will move 60% of the time is doing well. If a trader has an outstanding track record (as Robert Citron did in the early 1990s), it is likely to be a result of luck rather than superior trading skill. As our discussion of mutual fund performance in Chapter 4 shows, it appears that fund managers usually produce superior returns as a result of luck rather than skill.

Suppose that a financial institution employs 16 traders and one of those traders makes profits in every quarter of a year. Should the trader receive a good bonus? Should the trader's risk limits be increased? The answer to the first question is that the trader will inevitably receive a good bonus. The answer to the second question should be no. The chance of making a profit in four consecutive quarters from random trading is 0.5^4 or 1 in 16. This means that just by chance one of the 16 traders will "get it right" every single quarter of the year. We should not assume that the trader's luck will continue and we should not increase the trader's risk limits.

30.1.3 Do Not Underestimate the Benefits of Diversification

When a trader appears good at predicting a particular market variable, there is a tendency to increase the trader's risk limits. We have just argued that this is a bad idea because it is quite likely that the trader has been lucky rather than clever. However, let us suppose that we are really convinced that the trader has special talents. How undiversified should we allow ourselves to become in order to take advantage of the trader's special skills? As indicated in Section 1.1, the benefits from diversification are large. A trader has to be very good for it to be worth foregoing these benefits to speculate heavily on just one market variable.

An example will illustrate the point here. Suppose that there are 20 stocks, each of which has an expected return of 10% per annum and a standard deviation of return of

30%. The correlation between the returns from any two of the stocks is 0.2. By dividing an investment equally among the 20 stocks, an investor has an expected return of 10% per annum and standard deviation of returns of 14.7%. Diversification enables the investor to reduce risks by over half. Another way of expressing this is that diversification enables an investor to double the expected return per unit of risk taken. The investor would have to be very good at stock picking to consistently get a better risk–return trade-off by choosing undiversified stock portfolios.

30.1.4 Carry Out Scenario Analyses and Stress Tests

As discussed in Chapter 16, the calculation of risk measures such as VaR should always be accompanied by scenario analyses and stress testing to obtain an understanding of what can go wrong. Without the discipline of stress testing, human beings have an unfortunate tendency to anchor on one or two scenarios when evaluating decisions. In 1993 and 1994, for example, Procter & Gamble was so convinced that interest rates would remain low that it ignored the possibility of a 100-basis-point increase in its decision making.

Once stress-testing results have been produced, they should become inputs to a financial institution's strategic decision making. All too often, particularly when times are good, the results of stress testing are ignored. This happened at a number of financial institutions prior to July 2007.

30.2 Managing the Trading Room

In trading rooms there is a tendency to regard high-performing traders as “untouchable” and to not subject their activities to the same scrutiny as other traders. Apparently Joseph Jett, Kidder Peabody's star trader of Treasury instruments, was often “too busy” to answer questions and discuss his positions with the company's risk managers.

All traders—particularly those making high profits—should be fully accountable. It is important for the financial institution to know whether the high profits are being made by taking unreasonably high risks. It is also important to check that the financial institution's computer systems and pricing models are correct and are not being manipulated in some way.

30.2.1 Separate the Front, Middle, and Back Office

The *front office* in a financial institution consists of the traders who are executing trades, taking positions, and so on. The *middle office* consists of risk managers who are monitoring the risks being taken. The *back office* is where the record keeping and accounting takes place. Some of the worst derivatives disasters have occurred because these functions were not kept separate. Nick Leeson controlled both the front and back office for Barings in Singapore and was, as a result, able to conceal the disastrous nature of his trades from his superiors in London for some time.

30.2.2 Do Not Blindly Trust Models

We discussed model risk in Chapter 22. Some of the large losses experienced by financial institutions arose because of the models and computer systems being used. Kidder Peabody was misled by its own systems. Another example of an incorrect model leading to losses is provided by National Westminster Bank. This bank had an incorrect model for valuing swap options that led to a loss.

If large profits are reported when relatively simple trading strategies are followed, there is a good chance that the models underlying the calculation of the profits are wrong. Similarly, if a financial institution appears to be particularly competitive on its quotes for a particular type of deal, there is a good chance that it is using a different model from other market participants—which almost certainly means that, before too long, it will have to change its model and report a loss. Getting too much business of a certain type can be just as worrisome to the head of a trading room as getting too little business of that type.

30.2.3 Be Conservative in Recognizing Inception Profits

When a financial institution sells a highly exotic instrument to a nonfinancial corporation, the valuation can be highly dependent on the underlying model. For example, instruments with long-dated embedded interest rate options can be highly dependent on the interest rate model used. In these circumstances, a phrase used to describe the daily marking to market of the deal is *marking to model*. This is because there are no market prices for similar deals that can be used as a benchmark.

Suppose that a financial institution manages to sell an instrument to a client for \$10 million more than it is worth—or at least \$10 million more than its model says it is worth. The \$10 million is known as an *inception profit*. When should it be recognized? There appears to be a lot of variation in what different derivatives dealers do. Some recognize the \$10 million immediately, whereas others are much more conservative and recognize it slowly over the life of the deal.

Recognizing inception profits immediately is very dangerous. It encourages traders to use aggressive models, take their bonuses, and leave before the model and the value of the deal come under close scrutiny. It is much better to take reserves and to recognize inception profits slowly, so that traders are motivated to investigate the impact of several different models and several different sets of assumptions before committing themselves to a deal.

30.2.4 Do Not Sell Clients Inappropriate Products

It is tempting to sell corporate clients inappropriate products, particularly when they appear to have an appetite for the underlying risks. But this is shortsighted. A dramatic illustration is provided by the activities of Bankers Trust (BT) in the period leading up to the spring of 1994. BT's clients, such as Procter & Gamble, were persuaded to buy high-risk and totally inappropriate products. A typical product would give a client a

good chance of saving a few basis points on its borrowings and a small chance of losing a large amount of money. The products worked well for BT's clients in 1992 and 1993, but blew up in 1994 when interest rates rose sharply. The bad publicity that followed hurt BT greatly. The years it had spent building up trust among corporate clients and developing an enviable reputation for innovation in derivatives were largely lost as a result of the activities of a few overly aggressive salesmen. BT was forced to pay large amounts of money to its clients to settle lawsuits out of court. It was taken over by Deutsche Bank in 1999.

30.2.5 Beware of Easy Profits

Enron provides an example of how overly aggressive deal makers can cost their banks billions of dollars. Doing business with Enron seemed very profitable and banks competed with each other for this business. But the fact that many banks push hard to get a certain type of business should not be taken as an indication that the business will ultimately be profitable. The business that Enron did with banks resulted in many very expensive shareholder lawsuits. In general, transactions where high profits seem easy to achieve should be looked at closely for potential operational, credit, or market risks.

Investing in the AAA-rated tranches of ABSs and ABS CDOs that were created from subprime mortgages (see Chapter 7) seemed like a money-making machine for many banks. The promised returns were higher than the returns normally earned on AAA-rated instruments. Most traders did not stop to ask whether this was because there were risks that had not been taken into account by rating agencies.

30.3 Liquidity Risk

We discussed liquidity risk in Chapter 21. Financial engineers usually base the pricing of exotic instruments and other instruments that trade relatively infrequently on the prices of actively traded instruments. For example:

1. A trader often calculates a zero curve from actively traded government bonds (known as on-the-run bonds) and uses it to price bonds that trade less frequently (off-the-run bonds).
2. A trader often implies the volatility of an asset from actively traded options and uses it to price less actively traded options.
3. A trader often implies information about the behavior of interest rates from actively traded interest rate caps and swap options and uses it to price products that are highly structured.

These practices are not unreasonable. However, it is dangerous to assume that less actively traded instruments can always be traded at close to their theoretical prices. When financial markets experience a shock of one sort or another, liquidity black holes may develop (see Section 21.3). Liquidity then becomes very important to investors, and illiquid instruments often sell at a big discount to their theoretical values.

An example of liquidity risk is provided by Long-Term Capital Management (LTCM), which is the subject of Business Snapshot 16.1. This hedge fund followed a strategy known as *convergence arbitrage*. It attempted to identify two securities (or portfolios of securities) that should in theory sell for the same price. If the market price of one security was less than that of the other, it would buy that security and sell the other. The strategy is based on the idea that if two securities have the same theoretical price, their market prices should eventually be the same.

In the summer of 1998, LTCM took a huge loss. This was largely because a default by Russia on its debt caused a flight to quality. LTCM tended to be long illiquid instruments and short the corresponding liquid instruments. (For example, it was long off-the-run bonds and short on-the-run bonds.) The spreads between the prices of illiquid instruments and the corresponding liquid instruments widened sharply after the Russian default. LTCM was highly leveraged. It experienced huge losses and there were margin calls on its positions that it had difficulty meeting. The Archegos story also emphasizes the dangers of high leverage.

The LTCM and Archegos stories reinforce the importance of carrying out scenario analyses and stress testing to look at what can happen in extreme scenarios. As discussed in Chapter 16, it is important to consider not only immediate losses but also losses created by knock-on effects.

30.3.1 Beware When Everyone Is Following the Same Trading Strategy

It sometimes happens that many market participants are following essentially the same trading strategy. This creates a dangerous environment where there are liable to be big market moves, liquidity black holes, and large losses for the market participants.

We gave one example of this in Business Snapshot 21.4, which discussed portfolio insurance and the market crash of October 1987. In the months leading up to the crash, increasing numbers of portfolio managers were attempting to insure their portfolios by creating synthetic put options. This involved buying stocks or stock index futures after a rise in the market and selling them after a fall. The result was an unstable market. A relatively small decline in stock prices could lead to a wave of selling by portfolio insurers. The latter would lead to a further decline in the market, which could give rise to another wave of selling, and so on. There is little doubt that, without portfolio insurance, the crash of October 1987 would have been much less severe.

Another example is provided by LTCM in 1998. Its position was made more difficult by the fact that many other hedge funds were following similar convergence arbitrage strategies. After the Russian default and the flight to quality, LTCM tried to liquidate part of its portfolio to meet margin calls. Unfortunately, other hedge funds were facing similar problems to LTCM and trying to do similar trades. This exacerbated the situation, causing liquidity spreads to be even higher than they would otherwise have been and reinforcing the flight to quality. Consider, for example, LTCM's position in U.S. Treasury bonds. It was long the illiquid off-the-run bonds and short the liquid on-the-run bonds. When a flight to quality caused spreads between yields on the two types of bonds to

widen, LTCM had to liquidate its positions by selling off-the-run bonds and buying on-the-run bonds. Other large hedge funds were doing the same. As a result, the price of on-the-run bonds rose relative to off-the-run bonds and the spread between the two yields widened even more.

A further example is provided by British insurance companies in the late 1990s. This is discussed in Business Snapshot 3.1. All insurance companies decided to hedge their exposure to a fall in long-term rates at about the same time. The result was a fall in long-term rates!

The key lesson to be learned from these stories is that it is important to see the big picture of what is going on in financial markets and to understand the risks inherent in situations where many market participants are liable to follow the same trading strategy.

30.3.2 Do Not Make Excessive Use of Short-Term Funding for Long-Term Needs

All financial institutions finance long-term needs with short-term sources of funds to some extent. But a financial institution that relies too heavily on short-term funds is likely to expose itself to unacceptable liquidity risks.

During the period leading up to the credit crisis of 2007, there was a tendency for subprime mortgages and other long-term assets to be financed by commercial paper while they were in a portfolio waiting to be packaged into structured products. Conduits and special purpose vehicles had an ongoing requirement for this type of financing. The commercial paper would typically be rolled over every month. For example, the purchasers of commercial paper issued on April 1 would be redeemed with the proceeds of a new commercial paper issue on May 1; this new commercial paper issue would in turn be redeemed with another new commercial paper issue on June 1; and so on. When investors lost confidence in subprime mortgages in August 2007, it became impossible to roll over commercial paper. In many instances, banks had provided guarantees and had to provide financing. This led to a shortage of liquidity. As a result, the credit crisis was more severe than it would have been if longer-term financing had been arranged.

Many of the failures of financial institutions during the crisis (e.g., Lehman Brothers and Northern Rock) were caused by excessive reliance on short-term funding. Once the market (rightly or wrongly) becomes concerned about the health of a financial institution, it can be impossible to roll over the financial institution's short-term funding. The Basel Committee has recognized the importance of liquidity risks by introducing liquidity requirements in Basel III.

30.3.3 Market Transparency Is Important

One of the lessons from the credit crisis of 2007 is that market transparency is important. During the period leading up to 2007, investors had been trading highly structured products without any real knowledge of the underlying assets. All they knew was the credit rating of the security being traded. With hindsight, we can say that investors should have

demanded more information about the underlying assets and should have more carefully assessed the risks they were taking.

The subprime meltdown of August 2007 caused investors to lose confidence in all structured products and withdraw from that market. This led to a market breakdown where tranches of structured products could only be sold at prices well below their theoretical values. There was a flight to quality and credit spreads increased. If there had been market transparency, so that investors understood the asset-backed securities they were buying, there would still have been subprime losses, but the flight to quality and disruptions to the market would have been less pronounced.

30.4 Lessons for Nonfinancial Corporations

Here are some lessons applicable primarily to nonfinancial corporations.

30.4.1 Make Sure You Fully Understand the Trades You Are Doing

Corporations should never undertake a trade or a trading strategy that they do not fully understand. This is a somewhat obvious point, but it is surprising how often a trader working for a nonfinancial corporation will, after a big loss, admit to not really understanding what was going on and claim to have been misled by investment bankers. Robert Citron, the treasurer of Orange County did this. So did the traders working for Hammersmith and Fulham, who in spite of their huge positions were surprisingly uninformed about how the swaps and other interest rate derivatives they traded really worked.

If a senior manager in a corporation does not understand a trade proposed by a subordinate, the trade should not be approved. A simple rule of thumb is that if a trade and the rationale for entering into it are so complicated that it cannot be easily understood, it is almost certainly inappropriate for the corporation. The trades undertaken by Procter & Gamble would have been vetoed under this criterion.

One way of ensuring that you fully understand a financial instrument is to value it. If a corporation does not have the in-house capability to value an instrument, it should not trade it. In practice, corporations often rely on their investment bankers for valuation advice. This is dangerous, as Procter & Gamble found out. When it wanted to unwind its transactions, it found it was facing prices produced by Bankers Trust's proprietary models, which it had no way of checking.

30.4.2 Make Sure a Hedger Does Not Become a Speculator

One of the unfortunate facts of life is that hedging is relatively dull, whereas speculation is exciting. When a company hires a trader to manage foreign-exchange risk, commodity price risk, or interest rate risk, there is a danger that the following happens. At first the trader does the job diligently and earns the confidence of top management. He or she assesses the company's exposures and hedges them. As time goes by, the trader

becomes convinced that he or she can outguess the market. Slowly the trader becomes a speculator. At first things go well, but then a loss is made. To recover the loss, the trader doubles down. Further losses are made, and so on. The result is a disaster.

As mentioned earlier, clear limits to the risks that can be taken should be set by senior management. Controls should be put in place to ensure that the limits are obeyed. The trading strategy for a corporation should start with an analysis of the risks facing the corporation in foreign exchange, interest rate, and other markets. A decision should then be taken on how the risks are to be reduced to acceptable levels. It is a clear sign that something is wrong within a corporation if the trading strategy is not derived in a very direct way from the company's exposures.

30.4.3 Be Cautious about Making the Treasury Department a Profit Center

In the last 30 years, there has been a tendency to make the treasury department within a corporation a profit center. This seems to have much to recommend it. The treasurer is motivated to reduce financing costs and manage risks as profitably as possible. The problem is that the potential for the treasurer to make profits is limited. When raising funds and investing surplus cash, the treasurer is facing an efficient market. The treasurer can usually improve the bottom line only by taking additional risks. The company's hedging program gives the treasurer some scope for making shrewd decisions that increase profits. But it should be remembered that the goal of a hedging program is to reduce risks, not to increase expected profits. The decision to hedge will lead to a worse outcome than the decision not to hedge roughly 50% of the time. The danger of making the treasury department a profit center is that the treasurer is motivated to become a speculator. An Orange County or Procter & Gamble type of outcome is then liable to occur.

30.5 A Final Point

Most of the risks we have considered in this book are what are termed *known risks*. They are risks such as market risks and credit risks that can be quantified using historical data. Two other types of risk are important to financial institutions: *unknown risks* and *unknowable risks*.

Unknown risks are risks where the event that could cause a loss is known, but its probability of occurrence cannot easily be determined. Operational risks and business risks include many different types of unknown risks. What is the probability of a rogue trader loss? What is the probability of a loss from a major lawsuit? What is the probability that operations in a particular country will be expropriated by the government? These probabilities cannot usually be estimated using historical data. As discussed in Chapter 16, subjective probabilities are often used. In a widely referenced book, Knight (1921) uses the term "risk" to refer to known risks and the term "uncertainty" to refer to unknown risks.¹

¹See F. H. Knight, *Risk, Uncertainty and Profit* (Boston: Houghton Mifflin, 1921).

Unknowable risks are risks where even the event that could cause a loss is not known. Unknowable risks are in many ways the most insidious because they come as a complete surprise and often lead to dramatic losses. An unknowable risk is sometimes referred to as a *black swan*. (Black swans were not considered possible until they were discovered in Australia.) As pointed out by Taleb (2007), once it has occurred, a black swan event is often considered to be obvious.² Did the producers of multivolume encyclopedias in 1970 consider the possibility that technological developments would render their product worthless? Probably not, but ex post it seems a fairly obvious risk.

How can companies manage unknown and unknowable risks? A key tool is flexibility. Companies should avoid excessive leverage and try to ensure that their costs are variable rather than fixed as far as possible. Diversification across products and markets also increases flexibility. In the future, insurance companies may well develop more products to handle unknown and unknowable risks. As discussed in Chapter 20, products have already been developed to provide protection against some operational (unknown) risks. Handling unknowable risk is a challenging (but not totally impossible) contract design problem.

Further Reading

- Diebold, F. X., N. A. Doherty, and R. J. Herring. *The Known, the Unknown, and the Unknowable in Financial Risk Management*. Princeton, NJ: Princeton University Press, 2010.
- Dunbar, N. *Inventing Money: The Story of Long-Term Capital Management and the Legends Behind It*. Chichester, UK: John Wiley & Sons, 2000.
- Gomory, R. "The Known, the Unknown and the Unknowable." *Scientific American*, June 1995.
- Jorion, P. *Big Bets Gone Bad: Derivatives and Bankruptcy in Orange County*. New York: Academic Press, 1995.
- Jorion, P. "How Long-Term Lost Its Capital." *Risk* (September 1999): 31–36.
- Ju, X., and N. Pearson. "Using Value at Risk to Control Risk Taking: How Wrong Can You Be?" *Journal of Risk* 1 (1999): 5–36.
- Persaud, A. D., ed. *Liquidity Black Holes: Understanding Quantifying and Managing Financial Liquidity Risk*. London: Risk Books, 2003.
- Sorkin, A. R. *Too Big to Fail*. New York: Penguin, 2009.
- Tett, G. *Fool's Gold: How the Bold Dream of a Small Tribe at JPMorgan Was Corrupted by Wall Street Greed and Unleashed a Catastrophe*. New York: Free Press, 2009.
- Thomson, R. *Apocalypse Roulette: The Lethal World of Derivatives*. London: Macmillan, 1998.
- Zhang, P. G. *Barings Bankruptcy and Financial Derivatives*. Singapore: World Scientific Publishing, 1995.

²See N. N. Taleb, *The Black Swan: The Impact of the Highly Improbable* (New York: Random House, 2007).

Part Eight

APPENDICES

Appendix A

Compounding Frequencies for Interest Rates

A statement by a bank that the interest rate on one-year deposits is 10% per annum sounds straightforward and unambiguous. In fact, its precise meaning depends on the way the interest rate is measured.

If the interest rate is measured with annual compounding, the bank's statement that the interest rate is 10% means that \$100 grows to

$$\$100 \times 1.1 = \$110$$

at the end of one year. When the interest rate is measured with semiannual compounding, it means that we earn 5% every six months, with the interest being reinvested. In this case, \$100 grows to

$$\$100 \times 1.05 \times 1.05 = \$110.25$$

at the end of one year. When the interest rate is measured with quarterly compounding, the bank's statement means that we earn 2.5% every three months, with the interest being reinvested. The \$100 then grows to

$$\$100 \times 1.025^4 = \$110.38$$

at the end of one year. Table A.1 shows the effect of increasing the compounding frequency further.

The compounding frequency defines the units in which an interest rate is measured. A rate expressed with one compounding frequency can be converted into an equivalent

Table A.1 Effect of the Compounding Frequency on the Value of \$100 at the End of One Year When the Interest Rate is 10% per Annum

Compounding Frequency	Value of \$100 at End of Year (\$)
Annually ($m = 1$)	110.00
Semiannually ($m = 2$)	110.25
Quarterly ($m = 4$)	110.38
Monthly ($m = 12$)	110.47
Weekly ($m = 52$)	110.51
Daily ($m = 365$)	110.52

rate with a different compounding frequency. For example, from Table A.1 we see that 10.25% with annual compounding is equivalent to 10% with semiannual compounding. We can think of the difference between one compounding frequency and another to be analogous to the difference between kilometers and miles. They are two different units of measurement.

To generalize our results, suppose that an amount A is invested for n years at an interest rate of R per annum. If the rate is compounded once per annum, the terminal value of the investment is

$$A(1 + R)^n$$

If the rate is compounded m times per annum, the terminal value of the investment is

$$A\left(1 + \frac{R}{m}\right)^{mn} \quad (\text{A.1})$$

When $m = 1$ the rate is sometimes referred to as the *equivalent annual interest rate*.

A.1 Continuous Compounding

The limit as the compounding frequency, m , tends to infinity is known as *continuous compounding*.¹ With continuous compounding, it can be shown that an amount A invested for n years at rate R grows to

$$Ae^{Rn} \quad (\text{A.2})$$

where $e = 2.71828$. The function e^x , which is also written $\exp(x)$, is built into most calculators, so the computation of the expression in equation (A.2) presents no problems. In the example in Table A.1, $A = 100$, $n = 1$, and $R = 0.1$, so that the value to which A grows with continuous compounding is

$$100e^{0.1} = \$110.52$$

This is (to two decimal places) the same as the value with daily compounding. For most practical purposes, continuous compounding can be thought of as being equivalent to

¹Actuaries sometimes refer to a continuously compounded rate as the *force of interest*.

daily compounding. Compounding a sum of money at a continuously compounded rate R for n years involves multiplying it by e^{Rn} . Discounting it at a continuously compounded rate R for n years involves multiplying by e^{-Rn} .

Suppose that R_c is a rate of interest with continuous compounding and R_m is the equivalent rate with compounding m times per annum. From the results in equations (A.1) and (A.2), we have

$$Ae^{R_c n} = A \left(1 + \frac{R_m}{m} \right)^{mn}$$

or

$$e^{R_c} = \left(1 + \frac{R_m}{m} \right)^m$$

This means that

$$R_c = m \ln \left(1 + \frac{R_m}{m} \right) \quad (\text{A.3})$$

and

$$R_m = m(e^{R_c/m} - 1) \quad (\text{A.4})$$

These equations can be used to convert a rate with a compounding frequency of m times per annum to a continuously compounded rate and vice versa. The function $\ln$ is the natural logarithm function and is built into most calculators. It is defined so that if $y = \ln x$, then $x = e^y$.

Example A.1

Consider an interest rate that is quoted as 10% per annum with semiannual compounding. From equation (A.3), with $m = 2$ and $R_m = 0.1$, the equivalent rate with continuous compounding is

$$2 \ln \left(1 + \frac{0.1}{2} \right) = 0.09758$$

or 9.758% per annum.

Example A.2

Suppose that a lender quotes the interest rate on loans as 8% per annum with continuous compounding, and that interest is actually paid quarterly. From equation (A.4), with $m = 4$ and $R_c = 0.08$, the equivalent rate with quarterly compounding is

$$4(e^{0.08/4} - 1) = 0.0808$$

or 8.08% per annum. This means that on a \$1,000 loan, interest payments of \$20.20 would be required each quarter.

Appendix B

Zero Rates, Forward Rates, and Zero-Coupon Yield Curves

The n -year zero-coupon interest rate is the rate of interest earned on an investment that starts today and lasts for n years. All the interest and principal is realized at the end of n years. There are no intermediate payments. The n -year zero-coupon interest rate is sometimes also referred to as the n -year *spot interest rate*, the n -year *zero rate*, or just the n -year zero. The zero rate as a function of maturity is referred to as the *zero curve*. Suppose a five-year zero rate with continuous compounding is quoted as 5% per annum. (See Appendix A for a discussion of compounding frequencies.) This means that \$100, if invested for five years, grows to

$$100 \times e^{0.05 \times 5} = 128.40$$

A forward rate is the future zero rate implied by today's zero rates. Consider the zero rates shown in Table B.1. The forward rate for the period between six months and one year is 6.6%. This is because 5% for the first six months combined with 6.6% for the next six months gives an average of 5.8% for the first year. Similarly, the forward rate for the period between 12 months and 18 months is 7.6%, because this rate, when combined with 5.8% for the first 12 months, gives an average of 6.4% for the 18 months. In general, the forward rate F for the period between times T_1 and T_2 is

$$F = \frac{R_2 T_2 - R_1 T_1}{T_2 - T_1} \quad (\text{B.1})$$

Table B.1 Zero Rates

Maturity (years)	Zero Rate (%) (cont. comp.)
0.5	5.0
1.0	5.8
1.5	6.4
2.0	6.8

where R_1 is the zero rate for maturity T_1 and R_2 is the zero rate for maturity T_2 . This formula is exactly true when rates are measured with continuous compounding and approximately true for other compounding frequencies. The results from using this formula on the rates in Table B.1 are in Table B.2. For example, substituting $T_1 = 1.5$, $T_2 = 2.0$, $R_1 = 0.064$, and $R_2 = 0.068$ gives $F = 0.08$ showing that the forward rate for the period between 18 months and 24 months is 8%.

Investors who think that future interest rates will be markedly different from forward rates have no difficulty in finding trades that reflect their beliefs. Consider an investor who can borrow or lend at the rates in Table B.1. Suppose the investor thinks that the six-month interest rates will not change much over the next two years. The investor can borrow six-month funds and invest for two years. The six-month borrowings can be rolled over at the end of 6, 12, and 18 months. If interest rates do stay about the same, this strategy will yield a profit of about 1.8% per year because interest will be received at 6.8% and paid at 5%. This type of trading strategy is known as a *yield curve play*. The investor is speculating that rates in the future will be quite different from the forward rates shown in Table B.2.

Robert Citron, the treasurer at Orange County, used yield curve plays similar to the one we have just described very successfully in 1992 and 1993. The profit from Mr. Citron's trades became an important contributor to Orange County's budget and he was re-elected. In 1994, he used the same strategy more aggressively. If short-term interest rates had remained the same or declined, he would have done very well. As it happened, interest rates rose sharply during 1994. On December 1, 1994, Orange County announced that its investment portfolio had lost \$1.5 billion and several days later it filed for bankruptcy protection.

Table B.2 Forward Rates for Zero Rates in Table B.1

Period (years)	Forward Rate (%) (cont. comp.)
0.5 to 1.0	6.6
1.0 to 1.5	7.6
1.5 to 2.0	8.0

B.1 Bond Pricing

Most bonds provide coupons periodically. The bond's principal (which is also known as its par value or face value) is received at the end of its life. The theoretical price of a bond can be calculated as the present value of all the cash flows that will be received by the owner of the bond. The most accurate approach is to use a different zero rate for each cash flow. To illustrate this, consider the situation where zero rates are as in Table B.1. Suppose that a two-year bond with a principal of \$100 provides coupons at the rate of 6% per annum semiannually. To calculate the present value of the first coupon of \$3, we discount it at 5% for six months; to calculate the present value of the second coupon of \$3, we discount it at 5.8% for one year; and so on. The theoretical price of the bond is therefore:

$$3e^{-0.05 \times 0.5} + 3e^{-0.058 \times 1.0} + 3e^{-0.064 \times 1.5} + 103e^{-0.068 \times 2.0} = 98.39$$

or \$98.39.

B.2 Bond Yields

A bond's yield is the discount rate that, when applied to all cash flows, gives a bond price equal to its market price. Suppose that the theoretical price of the bond we have been considering, \$98.39, is also its market value (i.e., the market's price of the bond is in exact agreement with the data in Table B.1). If γ is the yield on the bond, expressed with continuous compounding, we must have:

$$3e^{-\gamma \times 0.5} + 3e^{-\gamma \times 1.0} + 3e^{-\gamma \times 1.5} + 103e^{-\gamma \times 2.0} = 98.39$$

This equation can be solved using Excel's Solver or in some other way to give $\gamma = 6.76\%$.

B.3 Treasury Rates

Treasury rates are the rates an investor earns on Treasury bills and Treasury bonds. These are the instruments used by a government to borrow in its own currency. Japanese Treasury rates are the rates at which the Japanese government borrows in yen; U.S. Treasury rates are the rates at which the U.S. government borrows in U.S. dollars; and so on.

B.4 Determining Treasury Zero Rates

One way of determining Treasury zero rates such as those in Table B.1 is to observe the yields on "strips." These are zero-coupon bonds that are artificially created by traders when they sell the coupons on a Treasury bond separately from the principal.

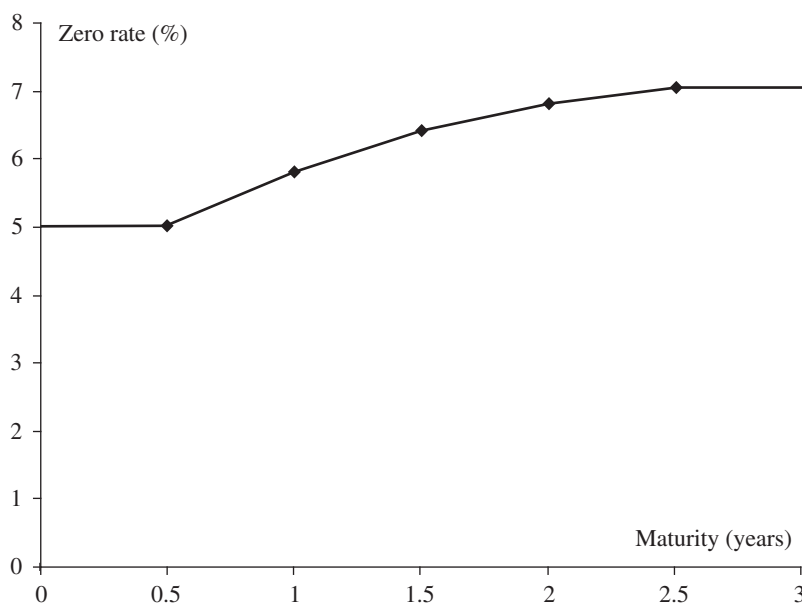
Table B.3 Rates After Two-Year Rate Has Been Determined Using the Bootstrap Method

Maturity (years)	Zero Rate (%) (cont. comp.)
0.5	5.00
1.0	5.80
1.5	6.40
2.0	6.80
2.5	??

Another way of determining Treasury zero rates is from regular Treasury bills and bonds. The most popular approach is known as the *bootstrap method*. This involves working from short maturities to successively longer maturities matching prices. Suppose that Table B.3 gives the Treasury rates determined so far and that a 2.5-year bond providing a coupon of 8% sells for \$102 per \$100 of principal. We would determine the 2.5-year zero-coupon interest rate as the rate which, when used in conjunction with the rates in Table B.3, gives the correct price for this bond. This involves solving

$$4e^{-0.05 \times 0.5} + 4e^{-0.058 \times 1.0} + 4e^{-0.064 \times 1.5} + 4e^{-0.068 \times 2.0} + 104e^{-R \times 2.5} = 102$$

which gives $R = 7.05\%$. The zero curve is usually assumed to be linear between the points that are determined by the bootstrap method. (In our example, the 2.25-year zero rate would be 6.9025%.) It is also assumed to be constant prior to the first point and beyond the last point. The zero curve for our example is shown in Figure B.1 .

**Figure B.1** Zero Curve for Data in Table B.3

Appendix C

Valuing Forward and Futures Contracts

The forward or futures price of an investment asset that provides no income is given by

$$F_0 = S_0 e^{rT}$$

where S_0 is the spot price of the asset today, T is the time to maturity of the forward or futures contract, and r is the continuously compounded risk-free rate for maturity T . When the asset provides income during the life of the contract that has a present value I , this becomes

$$F_0 = (S_0 - I)e^{rT}$$

When it provides a yield at rate q , it becomes

$$F_0 = S_0 e^{(r-q)T}$$

A foreign currency can be regarded as an investment asset that provides a yield equal to the foreign risk-free rate, so that the forward or futures price for a foreign currency is

$$F_0 = S_0 e^{(r-r_f)T}$$

where r_f is the foreign risk-free rate (continuously compounded) and S_0 is the spot exchange rate. The value of a forward contract where the holder has the right to buy the asset for a price of K is, in all cases,

$$(F_0 - K)e^{-rT}$$

where F_0 is the forward price, given by one of the formulas above. The value of a forward contract where the holder has the right to sell the asset for a price of K is similarly

$$(K - F_0)e^{-rT}$$

Example C.1

Consider a six-month futures contract on the S&P 500. The current value of the index is 1,200, the six-month risk-free rate is 5% per annum, and the average dividend yield on the S&P 500 over the next six months is expected to be 2% per annum (both rates continuously compounded). The futures price is $1,200e^{(0.05-0.02)\times 0.5}$ or 1,218.14.

Example C.2

The current forward price of a commodity for a contract maturing in nine months is \$550 per ounce. A company has a forward contract to buy 1,000 ounces of the commodity for a delivery price of \$530 in nine months. The nine-month risk-free rate is 4% per annum continuously compounded. The value of the forward contract is $1,000 \times (550 - 530)e^{-0.04 \times 9/12}$, or \$19,409.

In most situations forward and futures prices can be assumed to be the same. An exception is provided by contracts on interest rates. The quote for a Eurodollar futures contract should not be assumed to be the same as the corresponding forward interest rate, particularly when the contract has a long time to maturity.

Appendix D

Valuing Swaps

A plain vanilla interest rate swap can be valued by assuming that the interest rates that are realized in the future equal today's forward interest rates. As an example, consider an interest rate swap that has 14 months remaining and a notional principal of \$100 million. A fixed rate of 5% per annum is received and floating rate is paid, with exchanges taking place every six months. Assume that (a) four months ago, the six-month floating rate was 4%, (b) the forward floating interest rate for a six-month period starting in two months is 4.6%, and (c) the forward floating rate for a six-month period starting in eight months is 5.2%. All rates are expressed with semiannual compounding. The floating rate that will be exchanged in two months is equal the rate four months ago and is known. Assuming that forward rates are realized, the other two cash flows on the swap are as shown in Table D.1 (For example, in eight months the fixed-rate cash flow received is $0.5 \times 0.05 \times 100$, or \$2.5 million; the floating-rate cash flow paid is $0.5 \times 0.046 \times 100$, or \$2.3 million.) The value of the swap is the present value of the net cash flows in the final column.¹

D.1 Floating Forward Rates

Floating forward rates can be calculated from swap rates using a bootstrap method. Suppose that the 6-month, 12-month, 18-month, and 24-month swap rates (i.e., the fixed rates exchanged for the floating rate) are 3.8%, 4.3%, 4.6%, and 4.75% (with

¹Note that the precise timing of cash flows depends on day count conventions and holiday calendars.

Table D.1 Valuing an Interest Rate Swap by Assuming Forward Rates Are Realized

Time	Fixed Cash Flow (\$ million)	Floating Cash Flow (\$ million)	Net Cash Flow (\$ million)
2 months	2.5	−2.0	0.5
8 months	2.5	−2.3	0.2
14 months	2.5	−2.6	−0.1

continuous compounding). Suppose that in a two-year swap where cash flows are exchanged every six months the swap rate is 5%. Suppose further that the 0.5-year floating rate is 4%, the forward rate for the period between 6 months and 12 months is 5%, and the forward rate for the period between 12 months and 18 months is 5.5% (all semiannually compounded). The value of a two-year swap where floating is received and 5% is paid is zero. The value of the first exchange is

$$0.5 \times (0.04 - 0.05) \times 100 \times e^{-0.038 \times 0.5} = -0.4906$$

The value of the second exchange is

$$0.5 \times (0.05 - 0.05) \times 100 \times e^{-0.043 \times 1} = 0$$

The value of the third exchange is

$$0.5 \times (0.055 - 0.05) \times 100 \times e^{-0.046 \times 1.5} = 0.2333$$

The total value of the first three payments is $-0.4906 + 0 + 0.2333 = -0.2573$. Suppose that the (assumed unknown) forward rate for the final payment is F . For the swap to be worth zero, we must have

$$0.5 \times (F - 0.05) \times 100 \times e^{-0.0475 \times 2} = 0.2573$$

This gives $F = 0.05565$ or 5.565%.

D.2 Currency Swaps

A currency swap can be valued by assuming that exchange rates in the future equal today's forward exchange rates. As an example, consider a currency swap in which 4% will be received in GBP and 2% will be paid in USD once a year. The principals in the two currencies are 8 million USD and 5 million GBP. The swap will last for another three years. The swap cash flows are shown in the second and third columns of Table D.2. The forward exchange rates are (we assume) those shown in the fourth column. These are used to convert the GBP cash flows to USD. The final column shows the net cash flows. The value of the swap is the present value of these cash flows.

Table D.2 Valuing a Currency Swap by Assuming Forward Exchange Rates Are Realized (all cash flows in millions)

Time	USD Cash Flow	GBP Cash Flow	Forward Exchange Rate	USD Value of GBP Cash Flow	Net Cash Flow in USD
1	−0.16	0.2	1.2000	0.240	+0.080
2	−0.16	0.2	1.2400	0.248	+0.088
3	−0.16	0.2	1.2800	0.256	+0.096
3	−8.00	5.0	1.8800	6.400	−1.600

An alternative approach (which gives the same valuation) is to regard the swap as a long position in a GBP bond and a short position in a USD bond. Each bond can be valued in its own currency in the usual way and the current exchange rate can be used to convert the value of the GBP bond from GBP to USD.

Appendix E

Valuing European Options

The Black–Scholes–Merton formulas for valuing European call and put options on an investment asset that provides no income are

$$c = S_0 N(d_1) - Ke^{-rT} N(d_2)$$

and

$$p = Ke^{-rT} N(-d_2) - S_0 N(-d_1)$$

where

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$
$$d_2 = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}$$

The function $N(x)$ is the cumulative probability distribution function for a standardized normal distribution (see tables at the end of the book or Excel's NORMSDIST function). The variables c and p are the European call and European put price, S_0 is today's asset price, K is the strike price, r is the continuously compounded risk-free rate, σ is the stock price volatility, and T is the time to maturity of the option.

Table E.1 Greek Letters for Options on an Asset That Provides a Yield at Rate q

Greek Letter	Call Option	Put Option
Delta	$e^{-qT} N(d_1)$	$e^{-qT} [N(d_1) - 1]$
Gamma	$\frac{N'(d_1)e^{-qT}}{S_0 \sigma \sqrt{T}}$	$\frac{N'(d_1)e^{-qT}}{S_0 \sigma \sqrt{T}}$
Theta (per yr)	$-S_0 N'(d_1) \sigma e^{-qT} / (2 \sqrt{T})$ $+q S_0 N(d_1) e^{-qT}$ $-r K e^{-rT} N(d_2)$	$-S_0 N'(d_1) \sigma e^{-qT} / (2 \sqrt{T})$ $-q S_0 N(-d_1) e^{-qT}$ $+r K e^{-rT} N(-d_2)$
Vega (per %)	$\frac{S_0 \sqrt{T} N'(d_1) e^{-qT}}{100}$	$\frac{S_0 \sqrt{T} N'(d_1) e^{-qT}}{100}$
Rho (per %)	$\frac{K T e^{-rT} N(d_2)}{100}$	$-\frac{K T e^{-rT} N(-d_2)}{100}$

When the underlying asset provides a cash income, the present value of the income during the life of the option should be subtracted from S_0 . When the underlying asset provides a yield at rate q , the formulas become:

$$c = S_0 e^{-qT} N(d_1) - K e^{-rT} N(d_2)$$

and

$$p = K e^{-rT} N(-d_2) - S_0 e^{-qT} N(-d_1)$$

where

$$d_1 = \frac{\ln(S_0/K) + (r - q + \sigma^2/2)T}{\sigma \sqrt{T}}$$

$$d_2 = \frac{\ln(S_0/K) + (r - q - \sigma^2/2)T}{\sigma \sqrt{T}} = d_1 - \sigma \sqrt{T}$$

Options on a foreign currency can be valued by setting q equal to the foreign risk-free rate. Options on a futures or forward price can be valued using these formulas by setting S_0 equal to the current forward or futures price, $q = r$, and σ equal to the volatility of the forward or futures price.

Table E.1 gives formulas for the Greek letters for European options on an asset that provides income at rate q . (See Chapter 15 for definitions of and discussion about Greek letters.) $N'(x)$ is the standard normal density function:

$$N'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

The implied volatility is the volatility, σ , which when substituted into a Black–Scholes–Merton pricing formula gives the market price. (See Section 8.2.) When Greek letters are calculated, the usual practice is to set σ equal to the implied volatility.

Example E.1

Consider a six-month European call option on a stock index. The current value of the index is 1,200, the strike price is 1,250, the risk-free rate is 5%, the dividend yield on the index is 2%, and the index volatility is 20%. In this case, $S_0 = 1,200$, $K = 1,250$, $r = 0.05$, $q = 0.02$, $\sigma = 0.2$, and $T = 0.5$. The value of the option is 53.44, the delta of the option is 0.45, the gamma is 0.0023, the theta is -0.22 , the vega is 3.33, and rho is 2.44. Note that the formula in Table E.1 gives theta per year. The theta quoted here is per calendar day.

The calculations in this appendix can be done with the RMFI software, available on the author's website, by selecting Option Type: Black–Scholes European. Option valuation is described more fully in Hull (2022).¹

¹See J. C. Hull, *Options, Futures, and Other Derivatives*, 11th ed. (Pearson, 2022).

Appendix F

Valuing American Options

To value American-style options, we divide the life of the option into n time steps of length Δt . Suppose that the asset price at the beginning of a step is S . At the end of the time step it moves up to Su with probability p and down to Sd with probability $1 - p$. For an investment asset that provides no income the values of u , d , and p are given by

$$u = e^{\sigma\sqrt{\Delta t}} \quad d = \frac{1}{u} \quad p = \frac{a - d}{u - d}$$

with

$$a = e^{r\Delta t}$$

where r is the risk-free rate and σ is the volatility.

Figure F.1 shows the tree constructed for valuing a five-month American put option on a non-dividend-paying stock where the initial stock price is 50, the strike price is 50, the risk-free rate is 10%, and the volatility is 40%. In this case, there are five steps so that $\Delta t = 0.08333$, $u = 1.1224$, $d = 0.8909$, $a = 1.0084$, and $p = 0.5073$. The upper number at each node is the stock price and the lower number is the value of the option.

At the final nodes of the tree, the option price is its intrinsic value. For example, at node G, the option price is $50 - 35.36 = 14.64$. At earlier nodes, we first calculate a value assuming that the option is held for a further time period of length Δt and then check to see whether early exercise is optimal. Consider first node E. If the option is

At each node:
Upper value = Underlying asset price
Lower value = Option price
Shading indicates where option is exercised.

Strike price = 50
Discount factor per step = 0.9917
Time step, $dt = 0.0833$ years, 30.42 days
Growth factor per step, $a = 1.0084$
Probability of up move, $p = 0.5073$
Up-step size, $u = 1.1224$
Down-step size, $d = 0.8909$

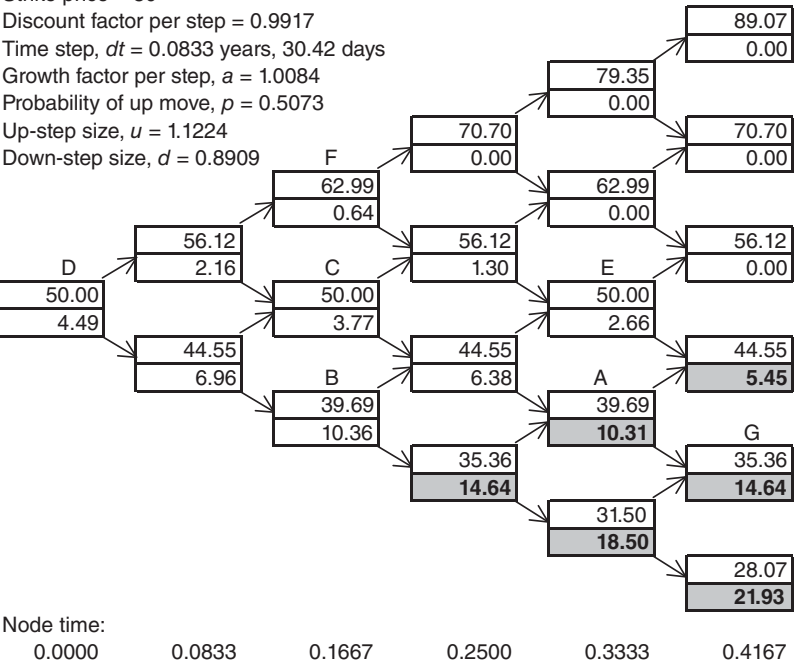


Figure F.1 Binomial Tree for American Put on Non-Dividend-Paying Stock

held for a further time period, it will be worth 0.00 if there is an up move (probability: p) and 5.45 if there is a down move (probability: $1 - p$). The expected value in time Δt is therefore $0.5073 \times 0 + 0.4927 \times 5.45$ or 2.686, and the 2.66 value at node E is calculated by discounting this at the risk-free rate of 10% for one month. The option should not be exercised at node E as the payoff from early exercise would be zero. Consider next node A. A calculation similar to that just given shows that, assuming it is held for a further time period, the option's value at node A is 9.90. If exercised, its value is $50 - 39.69 = 10.31$. In this case, it should be exercised and the value of being at node A is 10.31.

Continuing to work back from the end of the tree to the beginning, the value of the option at the initial node D is found to be 4.49. As the number of steps on the tree is increased, the accuracy of the option price increases. With 30, 50, and 100 time steps, we get values for the option of 4.263, 4.272, and 4.278, respectively.

To calculate delta, we consider the two nodes at time Δt . In our example, as we move from the lower node to the upper node, the option price changes from 6.96 to 2.16 and the stock price changes from 44.55 to 56.12. The estimate of delta is the change in the option price divided by the change in the stock price:

$$\text{Delta} = \frac{2.16 - 6.96}{56.12 - 44.55} = -0.41$$

To calculate gamma we consider the three nodes at time $2\Delta t$. The delta calculated from the upper two nodes (C and F) is -0.241 . This can be regarded as the delta for a stock price of $(62.99 + 50)/2 = 56.49$. The delta calculated from the lower two nodes (B and C) is -0.639 . This can be regarded as the delta for a stock price of $(50 + 39.69)/2 = 44.84$. The estimate of gamma is the change in delta divided by the change in the stock price:

$$\text{Gamma} = \frac{-0.241 - (-0.639)}{56.49 - 44.84} = 0.034$$

We estimate theta from nodes D and C as

$$\text{Theta} = \frac{3.77 - 4.49}{2 \times 0.08333}$$

or -4.30 per year. This is -0.0118 per calendar day. Vega is estimated by increasing the volatility, constructing a new tree, and observing the effect of the increased volatility on the option price. Rho is calculated similarly.

When the asset underlying the option provides a yield at rate q , the procedure is exactly the same except that $a = e^{(r-q)\Delta t}$ instead of $e^{r\Delta t}$ in the equation for p . (When the asset is a foreign currency q is set equal to the foreign risk-free rate.) When the option is on the forward or futures price, a is set equal to one and the tree shows the forward or futures price at each node. The calculations we have described can be done using the RMFI software, available on the author's website, by selecting Option Type: Binomial American. Binomial trees and other numerical procedures are described more fully in Hull (2022).¹

¹See J. C. Hull, *Options, Futures, and Other Derivatives*, 11th ed. (Pearson, 2022).

Appendix G

Taylor Series Expansions

Consider a function $z = F(x)$. When a small change Δx is made to x , there is a corresponding small change Δz in z . A first approximation to the relationship between Δz and Δx is

$$\Delta z = \frac{dz}{dx} \Delta x \quad (\text{G.1})$$

This relationship is exact if z is a linear function of x and approximate in other situations. A more accurate approximation is

$$\Delta z = \frac{dz}{dx} \Delta x + \frac{1}{2} \frac{d^2 z}{dx^2} (\Delta x)^2 \quad (\text{G.2})$$

This relationship is exact if z is a quadratic function of x and approximate in other situations. By adding more terms in the series, we can increase accuracy. The full expansion is

$$\Delta z = \frac{dz}{dx} \Delta x + \frac{1}{2!} \frac{d^2 z}{dx^2} (\Delta x)^2 + \frac{1}{3!} \frac{d^3 z}{dx^3} (\Delta x)^3 + \frac{1}{4!} \frac{d^4 z}{dx^4} (\Delta x)^4 + \cdots$$

where $!$ denotes the factorial function: $2! = 2 \times 1 = 2$; $3! = 3 \times 2 \times 1 = 6$; $4! = 4 \times 3 \times 2 \times 1 = 24$; and so on.

Example G.1

Consider the function $z = \sqrt{x}$ so that

$$\frac{dz}{dx} = \frac{1}{2x^{1/2}} \quad \frac{d^2z}{dx^2} = -\frac{1}{4x^{3/2}} \quad \frac{d^3z}{dx^3} = \frac{3}{8x^{5/2}}$$

Suppose that $x = 2$ and $\Delta x = 0.1$ so that $\Delta z = \sqrt{2.1} - \sqrt{2} = 0.034924$. When $x = 2$,

$$\frac{dz}{dx} = 0.35355 \quad \frac{d^2z}{dx^2} = -0.08839 \quad \frac{d^3z}{dx^3} = 0.06629$$

The first order approximation to Δz , given by equation (G.1), is

$$\Delta z = 0.35355 \times 0.1 = 0.035355$$

The second order approximation, given by equation (G.2), is

$$\Delta z = 0.35355 \times 0.1 + \frac{1}{2} \times (-0.08839) \times 0.1^2 = 0.034913$$

The third order approximation is

$$\Delta z = 0.35355 \times 0.1 + \frac{1}{2} \times (-0.08839) \times 0.1^2 + \frac{1}{6} \times 0.06629 \times 0.1^3 = 0.034924$$

It can be seen that the series expansion quickly converges to the correct answer of 0.034924.

G.1 Functions of Two Variables

Consider next a function of two variables, $z = F(x, y)$. Suppose that Δx and Δy are small changes in x and y , respectively, and that Δz is the corresponding small change in z . In this case, the first order approximation is

$$\Delta z = \frac{\partial z}{\partial x} \Delta x + \frac{\partial z}{\partial y} \Delta y \quad (\text{G.3})$$

The second order approximation is

$$\Delta z = \frac{\partial z}{\partial x} \Delta x + \frac{\partial z}{\partial y} \Delta y + \frac{1}{2} \frac{\partial^2 z}{\partial x^2} (\Delta x)^2 + \frac{1}{2} \frac{\partial^2 z}{\partial y^2} (\Delta y)^2 + \frac{\partial^2 z}{\partial x \partial y} (\Delta x \Delta y) \quad (\text{G.4})$$

Example G.2

Consider the function $z = \sqrt{xy}$ so that

$$\begin{aligned}\frac{\partial z}{\partial x} &= \frac{y^{1/2}}{2x^{1/2}} & \frac{\partial z}{\partial y} &= \frac{x^{1/2}}{2y^{1/2}} \\ \frac{\partial^2 z}{\partial x^2} &= -\frac{y^{1/2}}{4x^{3/2}} & \frac{\partial^2 z}{\partial y^2} &= -\frac{x^{1/2}}{4y^{3/2}} & \frac{\partial^2 z}{\partial x \partial y} &= \frac{1}{4(xy)^{1/2}}\end{aligned}$$

Suppose that $x = 2$, $y = 1$, $\Delta x = 0.1$, and $\Delta y = 0.1$ so that

$$\Delta z = \sqrt{2.1 \times 1.1} - \sqrt{2 \times 1} = 0.10565$$

When $x = 2$ and $y = 1$

$$\begin{aligned}\frac{\partial z}{\partial x} &= 0.35355 & \frac{\partial z}{\partial y} &= 0.70711 \\ \frac{\partial^2 z}{\partial x^2} &= -0.08839 & \frac{\partial^2 z}{\partial y^2} &= -0.35355 & \frac{\partial^2 z}{\partial x \partial y} &= 0.17678\end{aligned}$$

The first order approximation to Δz , given by equation (G.3), is

$$\Delta z = 0.35355 \times 0.1 + 0.70711 \times 0.1 = 0.10607$$

The second order approximation, given by equation (G.4), is

$$\begin{aligned}\Delta z &= 0.35355 \times 0.1 + 0.70711 \times 0.1 \\ &+ \frac{1}{2} \times (-0.08839) \times (0.1)^2 + \frac{1}{2} \times (-0.35355) \times (0.1)^2 \\ &+ 0.17678 \times 0.1 \times 0.1 = 0.10562\end{aligned}$$

The series solution is converging to the correct answer of 0.10565.

G.2 General Result

For a function z of n variables, $x_1, x_2, \dots, x_n$, the second order approximation given by a Taylor series for Δz is

$$\Delta z = \sum_{i=1}^n \frac{\partial z}{\partial x_i} \Delta x_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 z}{\partial x_i \partial x_j} \Delta x_i \Delta x_j$$

Appendix H

Eigenvectors and Eigenvalues

Consider an $n \times n$ matrix A and suppose that $\mathbf{x}$ is an $n \times 1$ vector. Consider the equation

$$A\mathbf{x} = \lambda\mathbf{x} \quad (\text{H.1})$$

The equation can be written

$$(A - \lambda I)\mathbf{x} = 0$$

where I is the $n \times n$ identity matrix (which is the $n \times n$ matrix with diagonal elements equal to 1 and all other elements equal to zero). Clearly, $\mathbf{x} = 0$ is a solution to equation (H.1). Under what circumstances are there other solutions? A theorem in linear algebra tells us that there are other solutions when the determinant of $A - \lambda I$ is zero. The values of λ that lead to solutions of equation (H.1) are therefore the values of λ that we get when we solve the equation that sets the determinant of $A - \lambda I$ equal to zero. This equation is an n th order polynomial in λ . In general, it has n solutions. The solutions are the *eigenvalues* of the matrix, A . The vector $\mathbf{x}$ that solves equation (H.1) for a particular eigenvalue is an *eigenvector*. In general, there are n eigenvectors, one corresponding to each eigenvalue.

As a simple example, suppose that

$$A = \begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix}$$

In this case

$$A - \lambda I = \begin{pmatrix} 1 - \lambda & -1 \\ 2 & 4 - \lambda \end{pmatrix}$$

The determinant of this matrix is

$$(1 - \lambda)(4 - \lambda) - (-1) \times 2 = \lambda^2 - 5\lambda + 6$$

The solutions to this equation are $\lambda = 3$ and $\lambda = 2$. These are the two eigenvalues of the matrix.

To determine the eigenvectors corresponding to $\lambda = 3$, we solve equation (H.1):

$$\begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix} \mathbf{x} = 3\mathbf{x}$$

Setting

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

the equation to be solved becomes

$$\begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 3 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

The simultaneous equations corresponding to this are

$$x_1 - x_2 = 3x_1$$

and

$$2x_1 + 4x_2 = 3x_2$$

Both these equations are equivalent to

$$x_2 + 2x_1 = 0$$

It follows that any pair of numbers x_1 and x_2 solve the equation when $\lambda = 3$ providing $x_2 = -2x_1$. By convention, the values of x_1 and x_2 that are reported are those for which the length of the vector $\mathbf{x}$ is 1. This means that $x_1^2 + x_2^2 = 1$. In this case, the solution where $\mathbf{x}$ has a length of 1 is $x_1 = \sqrt{0.2} = 0.447$ and $x_2 = -2\sqrt{0.2} = -0.894$. (An alternative is $x_1 = -0.447$ and $x_2 = 0.894$.) The solution

$$\mathbf{x} = \begin{pmatrix} 0.447 \\ -0.894 \end{pmatrix}$$

is the eigenvector corresponding to the first eigenvalue ($\lambda = 3$).

A similar calculation shows that equation (H.1) is satisfied for $\lambda = 2$ when $x_1 + x_2 = 0$. The solution where $\mathbf{x}$ has a length of 1 is $x_1 = \sqrt{0.5} = 0.707$ and $x_2 = -\sqrt{0.5} = -0.707$. (An alternative is $x_1 = -0.707$ and $x_2 = 0.707$.) The eigenvector corresponding to the second eigenvalue ($\lambda = 2$) is therefore

$$\mathbf{x} = \begin{pmatrix} 0.707 \\ -0.707 \end{pmatrix}$$

For matrices of larger size, a numerical procedure must be used to determine eigenvectors and eigenvalues. One such numerical procedure is provided by Press et al. (2007).¹

Applications of eigenvalues and eigenvectors are in Appendices I and J. Software for calculating eigenvectors and eigenvalues is on the author's website. (See Principal Components Analysis.)

¹See W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. (Cambridge, UK: Cambridge University Press, 2007).

Appendix I

Principal Components Analysis

A principal components analysis is concerned with understanding the structure of data on n correlated variables. The aim of the analysis is to replace the n variables by a smaller number of uncorrelated variables. In the example in Section 14.3, there are eight variables. These are the daily changes in the 1-year, 2-year, 3-year, 4-year, 5-year, 7-year, 10-year, and 30-year swap rates.

The first step in the analysis is to calculate a covariance matrix from the data. As explained in Section 9.3, this is an $n \times n$ matrix where the (i, j) entry is the covariance between variable i and variable j . The entries on the diagonal (where $i = j$) are variances.

The next step is to calculate the eigenvalues and eigenvectors for this matrix (see Appendix H). The eigenvectors are chosen to have length 1. (As explained in Appendix H, this means that the sum of the squares of their elements is 1.) The eigenvector corresponding to the highest eigenvalue is the first principal component; the eigenvector corresponding to the second highest eigenvalue is the second principal component; and so on. The principal components for the example in Section 14.3 are shown in Table 14.3.

The eigenvalue for the i th principal component, as a percentage of the sum of all the eigenvalues, is the percentage of the overall variance explained by the i th principal component. The square root of the i th eigenvalue is the standard deviation of the i th factor score (see Table 14.4).

Software for carrying out a principal components analysis is on the author's website.

Appendix J

Manipulation of Credit Transition Matrices

Suppose that A is an $n \times n$ matrix of credit rating changes in one year. This is a matrix such as the one shown in Table 19.1. Assuming that rating changes in successive time periods are independent, the matrix of credit rating changes in m years is A^m . If m is an integer, this can be readily calculated using the normal rules for matrix multiplication.

Consider next the problem of calculating the transition matrix for $1/m$ years where m is an integer. (For example, when we are interested in one-month changes, $m = 12$.) This is a more complicated problem because we need to calculate the m th root of a matrix. We first calculate eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ and the corresponding eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ of the matrix A . These are explained in Appendix H. They have the property that

$$A\mathbf{x}_i = \lambda_i\mathbf{x}_i \quad (\text{J.1})$$

Define X as an $n \times n$ matrix whose i th column is $\mathbf{x}_i$ and Λ as an $n \times n$ diagonal matrix (i.e., a matrix that has zero values everywhere except on the diagonal) where the i th diagonal element is λ_i . From equation (J.1), we have

$$AX = X\Lambda$$

so that

$$A = X\Lambda X^{-1}$$

Define Λ^* as a diagonal matrix where the i th diagonal element is $\lambda_i^{1/m}$. Then

$$\begin{aligned}(X\Lambda^*X^{-1})^m &= (X\Lambda^*X^{-1})(X\Lambda^*X^{-1})(X\Lambda^*X^{-1}) \dots (X\Lambda^*X^{-1}) = X(\Lambda^*)^mX^{-1} \\ &= X\Lambda X^{-1} = A\end{aligned}$$

showing that the m th root of A , and therefore the transition matrix for time $1/m$ years, is $X\Lambda^*X^{-1}$.

Some authors, such as Jarrow, Lando, and Turnbull (1997), prefer to handle this problem in terms of what is termed a *generator matrix*.¹ This is a matrix Γ such that the transition matrix for a short period of time, Δt , is $I + \Gamma\Delta t$ and the transition matrix for a longer period of time, t , is

$$\exp(t\Gamma) = \sum_{k=0}^{\infty} \frac{(t\Gamma)^k}{k!}$$

where I is the identity matrix (with ones on the diagonal and zeros elsewhere).

Software for manipulating credit transition matrices is on the author's website.

¹See R. A. Jarrow, D. Lando, and S. M. Turnbull, "A Markov Model for the Term Structure of Credit Spreads," *Review of Financial Studies* 10 (1997): 481–523.

Appendix K

Valuation of Credit Default Swaps

Credit default swaps (CDSs) are described in Chapter 17. They can be valued using (risk-neutral) default probability estimates.

Suppose that the hazard rate of a reference entity is 2% per annum for five years. Table K.1 shows survival probabilities and unconditional default probabilities (that is, default probabilities as seen at time zero) for each of the five years. From equation (17.2), the probability of survival to time t is $e^{-0.02t}$. The probability of default during a year is the probability of survival to the beginning of the year minus the probability of survival to the end of the year. For example, the probability of survival to time 2 years is $e^{-0.02 \times 2} = 0.9608$ and the probability of survival to time 3 years is $e^{-0.02 \times 3} = 0.9418$. The probability of default during the third year is $0.9608 - 0.9418 = 0.0190$.

We will assume that defaults always happen halfway through a year and that payments on a five-year credit default swap are made once a year, at the end of each year. We also assume that the risk-free interest rate is 5% per annum with continuous compounding and the recovery rate is 40%. There are three parts to the calculation. These are shown in Tables K.2, K.3, and K.4.

Table K.2 shows the calculation of the present value of the expected payments made on the CDS assuming that payments are made at the rate of s per year and the notional principal is \$1. For example, there is a 0.9418 probability that the third payment of s is made. The expected payment is therefore $0.9418s$ and its present value is $0.9418se^{-0.05 \times 3} = 0.8106s$. The total present value of the expected payments in Table K.2 is $4.0728s$.

Table K.1 Unconditional Default Probabilities and Survival Probabilities

Time (years)	Probability of Surviving to Year End	Probability of Default During Year
1	0.9802	0.0198
2	0.9608	0.0194
3	0.9418	0.0190
4	0.9321	0.0186
5	0.9048	0.0183

Table K.2 Calculation of the Present Value of Expected Payments (payment = s per annum)

Time (years)	Probability of Survival	Expected Payment	Discount Factor	PV of Expected Payment
1	0.9802	$0.9802s$	0.9512	$0.9324s$
2	0.9608	$0.9608s$	0.9048	$0.8694s$
3	0.9418	$0.9418s$	0.8607	$0.8106s$
4	0.9231	$0.9231s$	0.8187	$0.7558s$
5	0.9048	$0.9048s$	0.7788	$0.7047s$
Total				$4.0728s$

Table K.3 Calculation of the Present Value of Expected Payoff (notional principal = \$1)

Time (years)	Probability of Default	Recovery Rate	Expected Payoff (\$)	Discount Factor	PV of Expected Payoff (\$)
0.5	0.0198	0.4	0.0119	0.9753	0.0116
1.5	0.0194	0.4	0.0116	0.9277	0.0108
2.5	0.0190	0.4	0.0114	0.8825	0.0101
3.5	0.0186	0.4	0.0112	0.8395	0.0094
4.5	0.0183	0.4	0.0110	0.7985	0.0088
Total					0.0506

Table K.4 Calculation of the Present Value of Accrual Payment

Time (years)	Probability of Default	Expected Accrual Payment	Discount Factor	PV of Expected Accrual Payment
0.5	0.0198	$0.0099s$	0.9753	$0.0097s$
1.5	0.0194	$0.0097s$	0.9277	$0.0090s$
2.5	0.0190	$0.0095s$	0.8825	$0.0084s$
3.5	0.0186	$0.0093s$	0.8395	$0.0078s$
4.5	0.0183	$0.0091s$	0.7985	$0.0073s$
Total				$0.0422s$

Table K.3 shows the calculation of the present value of the expected payoff assuming a notional principal of \$1. As mentioned earlier, we are assuming that defaults always happen halfway through a year. For example, there is a 0.0190 probability of a payoff halfway through the third year. Given that the recovery rate is 40%, the expected payoff at this time is $0.0190 \times 0.6 \times 1 = 0.0114$. The present value of the expected payoff is $0.0114e^{-0.05 \times 2.5} = 0.0101$. The total present value of the expected payoffs in Table K.3 is \$0.0506.

As a final step, we evaluate in Table K.4 the accrual payment made in the event of a default. These accrual payments arise because the spread payments s are made in arrears so that, when a default occurs, a portion of a spread payment is owed. Consider the third year. There is a 0.0190 probability of a default halfway through the year. An accrual payment of $0.5s$ is due if there is a default. The expected accrual payment halfway through the third year is therefore $0.0190 \times 0.5s = 0.0095s$. Its present value is $0.0095se^{-0.05 \times 2.5} = 0.0084s$. The total present value of the expected accrual payments is $0.0422s$.

From Tables K.2 and K.4, the present value of the expected payments is

$$4.0728s + 0.0422s = 4.1150s$$

From Table K.3, the present value of the expected payoff is 0.0506. Equating the two, the CDS spread for a new CDS is given by

$$4.1150s = 0.0506$$

or $s = 0.0123$. The mid-market spread should be 0.0123 times the principal or 123 basis points per year. (This is roughly what we would expect from the relationship in equation (17.3): the recovery rate is 40% and the hazard rate is 2% so that equation (17.3) predicts a spread of $0.02 \times 0.6 = 0.012$.)

This example is designed to illustrate the calculation methodology. In practice, we are likely to find that the calculations are more extensive than those in Table K.2 to K.4 because (a) payments are often made more frequently than once a year and (b) we want to assume that defaults can happen more frequently than once a year.

K.1 Marking to Market a CDS

At the time it is negotiated, a CDS, like most other swaps, is worth close to zero. At later times it may have a positive or negative value. Suppose, for example, the credit default swap in our example had been negotiated some time ago for a spread of 150 basis points, the present value of the payments by the buyer would be $4.1150 \times 0.0150 = 0.0617$ and the present value of the payoff would be 0.0506, as given by Table K.3. The value of the swap to the seller would therefore be $0.0617 - 0.0506$ or 0.0111 times the principal. Similarly, the mark-to-market value of the swap to the buyer of protection would be -0.0111 times the principal.

The RMFI software that accompanies this book and can be downloaded from the author's website includes a worksheet for carrying out the above calculations.

Appendix L

Synthetic CDOs and Their Valuation

Synthetic collateralized debt obligations (CDOs) consist of tranches where one party (Party A) agrees to make payments to another party (Party B) that are equal to those losses on a specified portfolio of debt instruments that are in a certain range. In return, Party B agrees to make payments to Party A that are a certain proportion of the amount of principal that is being insured.

Suppose that the range of losses for a particular tranche is from α_L to α_H . The variables α_L and α_H are known as the attachment point and detachment point, respectively. If α_L is 8% and α_H is 18%, Party A pays to Party B the losses on the portfolio, as they are incurred, in the range 8% to 18% of the total principal of the portfolio. The first 8% of losses on the portfolio does not therefore affect the tranche. The tranche is responsible for the next 10% of losses and its notional principal (initially $18 - 8 = 10\%$ of the portfolio principal) reduces as these losses are incurred. The tranche is wiped out when losses exceed 18%. The payments that are made by Party B to Party A are made periodically at a specified rate applied to the remaining notional tranche principal. This specified rate is known as the *tranche spread*.

The usual assumption is that all the debt instruments in the portfolio have the same probability distribution for the time to default. Define $Q(t)$ as the probability of a debt instrument defaulting by time t . The one-factor Gaussian copula model of time to default

presented in Section 9.6 has become the standard market model for valuing a tranche of a collateralized debt obligation (CDO). From equation (9.13)

$$Q(t|F) = N \left\{ \frac{N^{-1}[Q(t)] - \sqrt{\rho}F}{\sqrt{1-\rho}} \right\} \quad (\text{L.1})$$

where $Q(t|F)$ is the probability of the i th entity defaulting by time t conditional on the value of the factor, F . In the calculation of $Q(t)$ it is usually assumed that the hazard rate for a company is constant. When a CDS spread or other credit spread is available, it can be used to determine the hazard rate using calculations similar to those in Appendix K in conjunction with a search procedure.

Suppose that the hazard rate is λ . Then:

$$Q(t) = 1 - e^{-\lambda t} \quad (\text{L.2})$$

From the properties of the binomial distribution, the probability of exactly k defaults by time t , conditional on F is

$$P(k, t|F) = \frac{n!}{(n-k)!k!} Q(t|F)^k [1 - Q(t|F)]^{n-k} \quad (\text{L.3})$$

Define

$$n_L = \frac{\alpha_L n}{1-R} \quad \text{and} \quad n_H = \frac{\alpha_H n}{1-R}$$

where R is the recovery rate (assumed constant). Also, define $m(x)$ as the smallest integer greater than x . The tranche suffers no losses when the number of defaults, k , is less than $m(n_L)$. It is wiped out when k is greater than or equal to $m(n_H)$. Otherwise the tranche principal at time t is a proportion

$$\frac{\alpha_H - k(1-R)/n}{\alpha_H - \alpha_L}$$

of the initial tranche principal. These results can be used in conjunction with equations (L.1), (L.2), and (L.3) to calculate the expected tranche principal at all times conditional on F . We can then integrate over F to find the (unconditional) expected tranche principal. This integration is usually accomplished with a procedure known as Gaussian quadrature. (The author's website provides the tools for integrating over a normal distribution using Gaussian quadrature.)

It is usually assumed that defaults happen at the midpoint of the intervals between payments. Similarly to Appendix K, we are interested in the following quantities:

1. The present value of the expected spread payments received by Party A,
2. The present value of the expected payments for tranche losses made by Party A, and
3. The present value of accrual payments received by Party A.

The spread payments received by Party A at a particular time are linearly dependent on the tranche principal at that time. The tranche loss payments made by Party A (assumed to be at the midpoint of an interval) is the change in the principal during the interval. The accrual payment received by Party A is proportional to the tranche loss payments. For any assumption about spreads, all three quantities of interest can therefore be calculated from the expected tranche principal on payment dates. The break-even spread can therefore be calculated analogously to the way it is calculated for CDSs in Appendix K.

Derivatives dealers calculate an implied copula correlation, ρ , from the spreads quoted in the market for tranches of CDOs and tend to quote these rather than the spreads themselves. This is similar to the practice in options markets of quoting Black–Scholes–Merton implied volatilities rather than dollar prices. Two correlations are used. The compound (tranche) correlation for a tranche is the correlation consistent with the tranche’s market value. The base correlation for $x\%$ is the correlation consistent with the market value of a tranche with attachment point at zero and detachment point at $x\%$. There is a correlation smile phenomenon in CDO markets similar to the volatility smile phenomenon in options markets.

The RMFI software that accompanies this book and can be downloaded from the author’s website includes a worksheet for carrying out the above calculations.¹

¹More details on the calculation can be found in J. C. Hull, *Options, Futures, and Other Derivatives*, 11th ed. (Pearson, 2022).

Appendix M

SIMM

As discussed in Chapter 6, some standard OTC derivatives transactions are cleared through CCPs and some continue to be cleared bilaterally. Regulators require initial and variation margin to be posted for bilaterally cleared transactions. The variation margin requires agreement on the valuation of outstanding transactions, and procedures have been established for resolving any disagreements on this. The calculation of initial margin is more complicated than valuing the transactions and there is more scope for different models to give different results. As a result there have been attempts to develop an industry standard.

Initial margin is specified in the regulations for portfolios of uncleared transactions between two parties as the gain in value over 10 days that we are 99% certain will not be exceeded in stressed market conditions. Note that initial margin is the mirror image of VaR. When we are calculating VaR, we are determining extreme percentiles of the loss distribution, but when we are calculating initial margin we are determining extreme percentiles of the gain distribution. This is because exposure increases as the uncollateralized value of a portfolio increases.

The Basel Committee proposed a grid approach for calculating initial margin, which specified initial margin as a percentage of notional principal for different types of transactions. This was unpopular because it did not incorporate netting. If a market participant entered into a certain transaction on Day 1 and an almost offsetting transaction on Day 5, both with the same counterparty, the initial margin on Day 5 would be almost double that on Day 1—even though the net exposure to the counterparty would be close to

zero. ISDA proposed what is known as the Standard Initial Margin Model (SIMM) as a way of overcoming this. This model has now been approved by regulators.¹

The contributions of delta and vega to the initial margin is handled using the weighted sensitivity approach in Section 13.5. They are each calculated as:

$$\sqrt{\sum_{i=1}^n \sum_{j=1}^n \delta_i \delta_j \rho_{ij} W_i W_j}$$

where the W_i is the risk weight for risk factor i (specified by the regulators), δ_i is the sensitivity of the position held to risk factor i (determined by the bank), and ρ_{ij} is the correlation between risk factors i and j (specified by the regulators). Because a 10-day time horizon with 99% confidence is used, a possible formula for W_i is

$$W_i = \sqrt{10} \times N^{-1}(0.99) \sigma_i \quad (\text{M.1})$$

where σ_i is the daily volatility (or standard deviation, in the case of interest rates, credit spreads, and volatilities) of the i th risk factor in stressed market conditions.

To calculate the incremental effect on initial margin of gamma risk, SIMM considers the situation where all deltas are zero and there is no cross gamma. From the results in Section 15.11.2 it can be seen that the mean and standard deviation of the change in the value of the portfolio over one day are:

$$\begin{aligned} E(\Delta P) &= \frac{1}{2} \sum_i \gamma_i \sigma_i^2 \\ SD(\Delta P) &= \sqrt{\frac{1}{2} \sum_{i,j} \rho_{ij}^2 \gamma_i \gamma_j \sigma_i^2 \sigma_j^2} \end{aligned} \quad (\text{M.2})$$

where γ_i is the gamma with respect to the i th risk factor.

Estimates of the mean and standard deviation of portfolio change over 10 days are obtained by replacing σ_i with $\sqrt{10} \sigma_i$. Defining

$$C_i = \frac{1}{2} \gamma_i (\sqrt{10} \sigma_i)^2$$

the mean, m , and standard deviation, s , of the 10-day change are therefore given by

$$\begin{aligned} m &= \sum_i C_i \\ s^2 &= 2 \sum_{i,j} \rho_{ij}^2 C_i C_j \end{aligned} \quad (\text{M.3})$$

¹See “ISDA SIMM: From Principles to Model Specification,” ISDA, March 3, 2016.

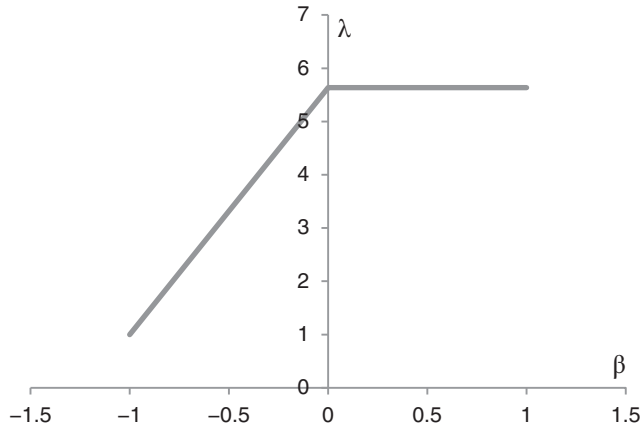


Figure M.1 Relation between λ and β

SIMM then sets

$$\text{IM}(\text{Gamma}) = m + \lambda_s / \sqrt{2}$$

The parameter λ in this equation is defined in terms of

$$\beta = \frac{\sum_i C_i}{\sum_i |C_i|}$$

as indicated in Figure M.1. This relationship produces results that have the right properties and correspond closely with tests carried out using Monte Carlo simulation.

There are a number of other details in SIMM. To simplify matters, gamma is calculated from vega using the relationship between the two that holds for European options. Risk factors are divided into buckets, and some risk factors involve term structures with vertices and are handled as we have described in Chapter 14. There are rules specified for calculating the correlations ρ_{ij} both within buckets and between buckets.

Answers to Questions and Problems

Chapter 1

1.1 The expected return is 12.5%. The standard deviation of returns is 17.07%.

1.2 From equations (1.1) and (1.2), expected return is 12.5%. SD of return is

$$\sqrt{0.5^2 \times 0.1707^2 + 0.5^2 \times 0.1707^2 + 2 \times 0.15 \times 0.5^2 \times 0.1707^2} = 0.1294$$

or 12.94%.

1.3	w_1	w_2	μ_p	σ_p ($\rho = 0.3$)	σ_p ($\rho = 1$)	σ_p ($\rho = -1$)
	0.0	1.0	15%	24.00%	24.00%	24.00%
	0.2	0.8	14%	20.39%	22.40%	16.00%
	0.4	0.6	13%	17.42%	20.80%	8.00%
	0.6	0.4	12%	15.48%	19.20%	0.00%
	0.8	0.2	11%	14.96%	17.60%	8.00%
	1.0	0.0	10%	16.00%	16.00%	16.00%

1.4 Nonsystematic risk can be diversified; systematic risk cannot. Systematic risk is more important to an equity investor. Either type of risk can lead to the bankruptcy of a corporation.

1.5 We assume that investors trade off mean return and standard deviation of return. For a given mean return, they want to minimize standard deviation of returns. All make the same estimates of means, standard deviations, and coefficients of

correlation for returns on individual investments. Furthermore, they can borrow or lend at the risk-free rate. The result is that they all want to be on the “new efficient frontier” in Figure 1.4. They choose the same portfolio of risky investments combined with borrowing or lending at the risk-free rate.

- 1.6** (a) 7.2%, (b) 9%, (c) 14.4%.
- 1.7** The capital asset pricing model assumes that there is one factor driving returns. Arbitrage pricing theory assumes multiple factors.
- 1.8** In many jurisdictions, interest on debt is deductible to the corporation whereas dividends are not deductible. It can therefore be more tax-efficient for a company to fund itself with debt. However, as debt increases, the probability of bankruptcy increases.
- 1.9** Risk decomposition refers to a procedure where risks are handled one by one. Risk aggregation refers to a procedure where a portfolio of risks is considered. Risk decomposition requires an in-depth understanding of individual risks. Risk aggregation also requires an understanding of the correlations between risks.
- 1.10** When potential losses are large, we cannot aggregate them and assume they will be diversified away. It is necessary to consider them one by one and handle them with insurance contracts, tighter internal controls, and the like.
- 1.11** This is the probability that profit is no worse than -4% of assets. This profit level is $4.6/1.5 = 3.067$ standard deviations from the mean. The probability that the bank will have a positive equity is therefore $N(3.067)$ where N is the cumulative normal distribution function. This is 99.89%.
- 1.12** Banks are allowed to accept deposits from the general public. Companies in retailing and manufacturing are not.
- 1.13** Professional fees (\$10 million per month), lost sales (people are reluctant to do business with a company that is being reorganized), and loss of key senior executives (lack of continuity).
- 1.14** The return earned by the hedge fund manager with zero alpha would be $0.05 + 0.6 \times (0.10 - 0.05) = 0.08$ or 8%. Because the alpha equals 4%, the hedge fund manager's return was 8% plus 4%, or 12%.

Chapter 2

- 2.1** The banking system became more concentrated, with large banks having a bigger share of the market. The total number of banks decreased from 14,483 to 4,796.
- 2.2** In the early twentieth century, many states passed laws restricting banks from opening more than one branch. The McFadden Act of 1927 restricted banks from opening branches in more than one state.
- 2.3** The main risk is that interest rates will rise so that, when deposits are rolled over, the bank has to pay a higher rate of interest. The result will be a reduction in the bank's net interest income. There is also liquidity risk.

-
- 2.4** DLC's loss is more than its equity capital, and it would probably be liquidated. The subordinated long-term debt holders would incur losses on their \$5 million investment. The depositors should get their money back.
- 2.5** The net interest income of a bank is interest received minus interest paid.
- 2.6** Credit risk primarily affects loan losses. Non-interest income includes trading gains and losses. Market risk therefore affects non-interest income. It also affects net interest income if assets and liabilities are not matched. Operational risk primarily affects non-interest expense.
- 2.7** A private placement is a new issue of securities that is sold to a small number of large institutional investors. A public offering is a new issue of securities that is offered to the general public. In a best efforts deal, the investment bank does as well as it can to place securities with investors, but does not guarantee that they can be sold. In a firm commitment deal, the investment bank agrees to buy the securities from the issuing company for a particular price and attempts to sell them in the market for a higher price.
- 2.8** The bidders when ranked from the highest price bid to the lowest are: H, C, F, A, B, D, E, and G. Bidders H, C, and F have bid for 140,000 shares. A has bid for 20,000. The price that clears the market is the price that was bid by A, or \$100. H, C, and F get their orders filled at this price. Half of A's order is filled at this price.
- 2.9** A Dutch auction potentially attracts a wide range of bidders. If all interested market participants bid, the price paid should be close to the market price immediately after the IPO. The usual IPO situation where the price turns out to be well below the market price should therefore be avoided. Also, investment banks are not able to restrict purchasers to their best current and potential clients. The Google IPO was different from a standard Dutch auction in that Google reserved the right to choose the number of shares that would be issued, and the percentage allocated to each bidder, when it saw the bids.
- 2.10** Poison pills can give management a negotiation tool, particularly if the board has the right to overturn a poison pill or make it ineffective. When it is confronted with a potential acquirer, the poison pill can buy the company time to bargain for a better purchase price or find other bidders. However, there is the danger that the poison pill will discourage potential buyers from approaching the company in the first place.
- 2.11** The brokerage subsidiary of a bank might recommend securities that the investment banking subsidiary is trying to sell. The commercial banking subsidiary might pass confidential information about its clients to the investment banking subsidiary. When a bank does business with a company (or wants to do business with the company), it might persuade the brokerage subsidiary to recommend the company's shares as a "buy." The commercial banking subsidiary might persuade a company to which it has lent money to do a bond issue because it is worried about its exposure to the client. (It wants the investment banking subsidiary to

persuade its clients to take on the credit risk.) These conflicts of interest are handled by what are known as Chinese walls. They prevent the flow of information from one part of the bank to another.

- 2.12** The trading book includes the instruments that a bank has as a result of its trading activities. The banking book consists of instruments it does not expect to trade, such as loans to corporations and individuals.
- 2.13** As a result of IFRS 9 and similar accounting standards from FASB, banks have to record loans they have made in their financial statements at the amount owed less an estimate of expected losses.
- 2.14** In the originate-to-distribute model, a bank originates loans and then securitizes them so that they are passed on to investors. This was done extensively with household mortgages during the seven-year period leading up to July 2007. In July 2007, investors lost confidence in the securitized products, and banks were forced to abandon the originate-to-distribute model, at least temporarily.

Chapter 3

- 3.1** Term life insurance lasts a fixed period (e.g., five years or ten years). The policyholder pays premiums. If the policyholder dies during the life of the policy, the policyholder's beneficiaries receive a payout equal to the principal amount of the policy. Whole life insurance lasts for the whole life of the policyholder. The policyholder pays premiums (usually the same each year), and the policyholder's beneficiaries receive a payout equal to the principal amount of the policy when the policyholder dies. There is an investment element to whole life insurance because the premiums in early years are high relative to the expected payout in those years. (The reverse is true in later years.)
- 3.2** Variable life insurance is whole life insurance where the policyholder can specify how the funds generated in early years (the excess of the premiums over the actuarial cost of the insurance) are invested. There is a minimum payout on death, but the payout can be more than the minimum if the investments do well. Universal life insurance is whole life insurance where the premium can be reduced to a specified minimum level without the policy lapsing. The insurance company chooses the investments (generally fixed income) and guarantees a minimum return. If the investments do well, the return provided on the policyholder's death may be greater than the guaranteed minimum.
- 3.3** Annuity contracts have exposure to longevity risk. Life insurance contracts have exposure to mortality risk.
- 3.4** The lifetime annuity created from an accumulated value was calculated using an interest rate that was the greater of (a) the market interest rate and (b) a prespecified minimum interest rate.

- 3.5** The probability that the woman will die during the first year is 0.002980. The probability that the woman will die during the second year is $0.003240 \times (1 - 0.002980) = 0.003230$. Suppose that the break-even premium is X . We must have

$$1,000,000 \times (0.002980 + 0.003230) = X + (1 - 0.002980)X$$

so that $X = 3,110$. The break-even premium is therefore \$3,110.

- 3.6** The probability of a male surviving to 30 is 0.97371. The probability of a male surviving to 90 is 0.18913. The probability of a male surviving to 90 conditional that 30 is reached is therefore $0.18913/0.97371 = 0.19424$. The probability of a female surviving to 90 conditional that 30 is reached is similarly $0.30509/0.98594 = 0.30944$.
- 3.7** The biggest risks are those arising from catastrophes such as earthquakes and hurricanes and those arising from liability insurance (e.g., claims related to asbestos in the United States). This is because there is no “law of large numbers” working in the insurance company’s favor. Either the event happens and there are big payouts or the event does not happen and there are no payouts.
- 3.8** CAT bonds (catastrophe bonds) are an alternative to reinsurance for an insurance company that has taken on a certain catastrophic risk (e.g., the risk of a hurricane or an earthquake) and wants to get rid of it. CAT bonds are issued by the insurance company. They provide a higher rate of interest than risk-free bonds. However, the bondholders agree to forgo interest, and possibly principal, to meet any claims against the insurance company that are within a prespecified range.
- 3.9** The CAT bond has very little systematic risk. Whether a particular type of catastrophe occurs is independent of the return on the market. The risks in the CAT bond are to some extent diversified away by the other investments in the portfolio. A B-rated bond does have systematic risk so that less of its risks are diversified away. It is likely therefore that the CAT bond is a better addition to the portfolio.
- 3.10** In Canada and the United Kingdom, health care is provided by the government. In the United States, publicly funded health care is limited and most individuals buy private health care insurance of one sort or another. In the United Kingdom, a private health care system operates alongside the public system.
- 3.11** Both moral hazard and adverse selection are potential problems. The insurance might lead to an individual not trying to keep a job as much as he or she otherwise would. Indeed, an individual might purposely lose his or her job to collect the insurance payout! Also, individuals who are most at risk for losing their jobs would be the ones who would choose to buy the insurance.
- 3.12** The payouts of property-casualty insurers show more variability than the payouts of life insurers. This is because of the possibility of catastrophes such as earthquakes and hurricanes and liability insurance claims such as those related to asbestos in the United States.
- 3.13** The loss ratio is the ratio of payouts to premiums in a year. The expense ratio is the ratio of expenses (e.g., sales commissions and expenses incurred in validating

losses) to premiums in a year. The statement is not true because investment income can be significant. Premiums are received at the beginning of a year, and payouts are made during the year or after the end of the year.

- 3.14** A defined contribution plan is a plan where the contributions of each employee (together with contributions made by the employer for that employee) are kept in a separate account and invested for the employee. When retirement age is reached, the accumulated amount is usually converted into an annuity. In a defined benefit plan, all contributions for all employees are pooled and invested. Employees receive a predefined pension that is based on their years of employment and final salary. At any given time, a defined benefit plan may be in surplus or in deficit.
- 3.15** The employee's wages are constant in real terms. Suppose that they are X per year. (The units for X do not matter for the purposes of our calculation.) The pension is $0.75X$. The real return earned is zero. Because employees work for 40 years, the present value of the contributions made by one employee is $40XR$ where R is the contribution rate as a percentage of the employee's wages. The present value of the benefits is $20 \times 0.75X = 15X$. The value of R that is necessary to adequately fund the plan must therefore satisfy

$$40XR = 15X$$

The solution to this equation is $R = 0.375$. The total of the employee contributions (and employer contributions, if any) should therefore be 37.5% of salary.

Chapter 4

- 4.1** The number of shares of an open-end mutual fund increases as investments in the fund increase and decreases as investors withdraw their funds. A closed-end fund is like any other corporation with a fixed number of shares that trade.
- 4.2** The net asset value (NAV) of an open-end mutual fund is calculated at 4 P.M. each day as the value of the assets held by the fund divided by the number of shares outstanding.
- 4.3** One NAV is the market value of the fund's shares, the other is the value of the fund's investments divided by the number of shares. The former is usually less than the latter.
- 4.4** An index fund is a fund that is designed so that its value tracks the performance of an index such as the S&P 500. It can be created by buying all the stocks (or a representative subset of the stocks) that underlie the index. Sometimes futures contracts on the index are used.
- 4.5** The front-end load is the amount an investor pays, as a percentage of his or her investment, when shares of the fund are purchased. The back-end load is the amount an investor pays, as a percentage of his or her investment, when shares of the fund are redeemed.

- 4.6** An exchange-traded fund (ETF) that tracks an index is created when an institutional investor deposits a portfolio of shares that is designed to track the index and receives shares in the ETF. Institutional investors can at any time exchange shares in the ETF for the underlying shares held by the ETF, or vice versa. The advantages over an open-end mutual fund that tracks the index are that the fund can be traded at any time, the fund can be shorted, and the fund does not have to be partially liquidated to accommodate redemptions. The advantage over a closed-end mutual fund is that there is very little difference between the ETF share price and the net asset value per share of the fund.
- 4.7** The arithmetic mean of a set of n numbers is the sum of the numbers divided by n . The geometric mean is the n th root of the product of the numbers. The arithmetic mean is always greater than or equal to the geometric mean. The return per year realized when an investment is held for several years is calculated using a geometric mean, not an arithmetic mean. (The procedure is to calculate the geometric mean of one plus the return in each year and then subtract one.)
- 4.8** Late trading is the illegal practice of putting in an order to buy or sell an open-end mutual fund at the 4 P.M. price after 4 P.M. Market timing is a practice where favored clients are allowed to buy and sell a mutual fund frequently to take advantage of the fact that some prices used in the calculation of the 4 P.M. net asset value are stale. Front running is the practice of trading by individuals ahead of a large institutional trade that is expected to move the market. Directed brokerage describes the situation where a mutual fund uses a brokerage house for trades when the brokerage house recommends the fund to clients.
- 4.9** Mutual funds must disclose their investment policies; their use of leverage is limited; they must calculate NAV daily; their shares must be redeemable at any time.
- 4.10** If a hedge fund is making money out of trading convertible bonds, it must be doing so at the expense of its counterparties. If most of the traders are hedge funds, they cannot all be making money.
- 4.11** Hurdle rate is the minimum return necessary for an incentive fee to be applicable. High-water mark refers to the previous losses that must be recouped before incentive fees are applicable. Clawback refers to investors being able to use some of the past incentive fees they have paid as an offset to current losses.
- 4.12** If the return is $X(> 2\%)$, the investors pay $0.02 + 0.2(X - 0.02)$ in fees. It must therefore be the case that

$$X - 0.02 - 0.2(X - 0.02) = 0.2$$

so that $0.8X = 0.216$ or $X = 0.27$. A return of 27% is necessary.

- 4.13** Short-term gains and losses do matter if the hedge fund is highly leveraged. Short-term losses can lead to margin calls that can destroy the hedge fund.
- 4.14** The leverage a hedge fund is allowed to take is limited by its prime broker. This in turn influences the risks that the hedge fund can take.

Chapter 5

- 5.1** When a trader enters into a long forward contract, she is agreeing to *buy* the underlying asset for a certain price at a certain time in the future. When a trader enters into a short forward contract, she is agreeing to *sell* the underlying asset for a certain price at a certain time in the future.
- 5.2** A trader is *hedging* when she has an exposure to the price of an asset and takes a position in a derivative to offset the exposure. In a *speculation*, the trader has no exposure to offset. She is betting on the future movements in the price of the asset. *Arbitrage* involves taking a position in two or more different markets to lock in a profit.
- 5.3** In the first case, the trader is obligated to buy the asset for \$50. (The trader does not have a choice.) In the second case, the trader has an option to buy the asset for \$50. (The trader does not have to exercise the option.)
- 5.4** Selling a call option involves giving someone else the right to buy an asset from you for a certain price. Buying a put option gives you the right to sell the asset to someone else.
- 5.5** (a) The investor is obligated to sell pounds for 1.3000 when they are worth 1.2900. The gain is $(1.3000 - 1.2900) \times 100,000 = \$1,000$.
(b) The investor is obligated to sell pounds for 1.3000 when they are worth 1.3200. The loss is $(1.3200 - 1.3000) \times 100,000 = \$2,000$.
- 5.6** (a) The trader sells for 50 cents per pound something that is worth 48.20 cents per pound. Gain = $(\$0.5000 - \$0.4820) \times 50,000 = \900 .
(b) The trader sells for 50 cents per pound something that is worth 51.30 cents per pound. Loss = $(\$0.5130 - \$0.5000) \times 50,000 = \650 .
- 5.7** You have sold a put option. You have agreed to buy 100 shares for \$40 per share if the party on the other side of the contract chooses to exercise the right to sell for this price. The option will be exercised only when the price of the stock is below \$40. Suppose, for example, that the option is exercised when the price is \$30. You have to buy at \$40 shares that are worth \$30; you lose \$10 per share, or \$1,000 in total. If the option is exercised when the price is \$20, you lose \$20 per share, or \$2,000 in total. The worst that can happen is that the price of the stock declines to almost zero during the three-month period. This highly unlikely event would cost you \$4,000. In return for the possible future losses, you receive the price of the option from the purchaser.
- 5.8** The over-the-counter (OTC) market is a market where financial institutions, fund managers, and corporate treasurers get in touch with each other directly and can enter into relatively large mutually acceptable transactions. An exchange-traded market is a market organized by an exchange where traders either meet physically or communicate electronically and the contracts that can be traded have been defined by the exchange. (a) OTC, (b) exchange, (c) both, (d) OTC, (e) OTC.

- 5.9** One strategy would be to buy 200 shares. Another would be to buy 2,000 options. If the share price does well, the second strategy will give rise to greater gains. For example, if the share price goes up to \$40, you gain $[2,000 \times (\$40 - \$30)] - \$5,800 = \$14,200$ from the second strategy and only $200 \times (\$40 - \$29) = \$2,200$ from the first strategy. However, if the share price does badly, the second strategy gives greater losses. For example, if the share price goes down to \$25, the first strategy leads to a loss of $200 \times (\$29 - \$25) = \$800$, whereas the second strategy leads to a loss of the whole \$5,800 investment. This example shows that options contain built-in leverage.
- 5.10** You could buy 5,000 put options (or 50 contracts) with a strike price of \$25 and an expiration date in four months. This provides a type of insurance. If, at the end of four months, the stock price proves to be less than \$25, you can exercise the options and sell the shares for \$25 each. The cost of this strategy is the price you pay for the put options.
- 5.11** A stock option provides no funds for the company. It is a security sold by one trader to another. The company is not involved. By contrast, a stock when it is first issued is a claim sold by the company to investors and does provide funds for the company.
- 5.12** Ignoring the time value of money, the holder of the option will make a profit if the stock price in March is greater than \$52.50. This is because the payoff to the holder of the option is, in these circumstances, greater than the \$2.50 paid for the option. The option will be exercised if the stock price at maturity is greater than \$50.00. Note that, if the stock price is between \$50.00 and \$52.50, and the option is exercised, the holder of the option takes a loss overall.
- 5.13** Ignoring the time value of money, the seller of the option will make a profit if the stock price in June is greater than \$56.00. This is because the cost to the seller of the option is in these circumstances less than the price received for the option. The option will be exercised if the stock price at maturity is less than \$60.00. Note that if the stock price is between \$56.00 and \$60.00, the seller of the option makes a profit even though the option is exercised.
- 5.14** A long position in a four-month put option can provide insurance against the exchange rate falling below the strike price. It ensures that the foreign currency can be sold for at least the strike price.
- 5.15** The company could enter into a long forward contract to buy 1 million Canadian dollars in six months. This would have the effect of locking in an exchange rate equal to the current forward exchange rate. Alternatively, the company could buy a call option giving it the right (but not the obligation) to purchase 1 million Canadian dollars at a certain exchange rate in six months. This would provide insurance against a strong Canadian dollar in six months while still allowing the company to benefit from a weak Canadian dollar at that time.
- 5.16** The payoff from an ICON is the payoff from:
- (a) A short position in call options to buy 169,000 yen with an exercise price of 1/169

- (b)** A long position in call options to buy 169,000 yen with an exercise price of 1/84.5

This is demonstrated by the following table:

	Terminal Value of Regular Bond	Terminal Value of Short Calls	Terminal Value of Long Calls	Terminal Value of Whole Position
$S_T > 169$	1,000	0	0	1,000
$84.5 \leq S_T \leq 169$	1,000	$-169,000 \left(\frac{1}{S_T} - \frac{1}{169} \right)$	0	$2,000 - \frac{169,000}{S_T}$
$S_T < 84.5$	1,000	$-169,000 \left(\frac{1}{S_T} - \frac{1}{169} \right)$	$169,000 \left(\frac{1}{S_T} - \frac{1}{84.5} \right)$	0

- 5.17 (a)** The trader buys a 180-day call option and takes a short position in a 180-day forward contract.
- (b)** The trader buys 90-day put options and takes a long position in a 90-day forward contract.
- 5.18** It enters into a five-year swap where it pays 3.30% and receives floating. Its investment is then at the floating rate minus 0.30%.
- 5.19** It enters into a five-year swap where it receives 3.26% and pays floating. Its net cost of borrowing is the floating rate plus 1.74%.
- 5.20** It enters into a three-year swap where it receives the floating rate and pays 3%. Its net borrowing cost for the three years is then 4% per annum.
- 5.21** Suppose that the weather is bad and the farmer's production is lower than expected. Other farmers are likely to have been affected similarly. Corn production overall will be low and as a consequence the price of corn will be relatively high. The farmer is likely to be overhedged relative to actual production. The farmer's problems arising from the bad harvest will be made worse by losses on the short futures position. This problem emphasizes the importance of looking at the big picture when hedging. The farmer is correct to question whether hedging price risk while ignoring other risks is a good strategy.
- 5.22** It may well be true that there is just as much chance that the price of oil in the future will be above the futures price as that it will be below the futures price. This means that the use of a futures contract for speculation would be like betting on whether a coin comes up heads or tails. But it might make sense for the airline to use futures for hedging rather than speculation. The futures contract then has the effect of reducing risks. It can be argued that an airline should not expose its shareholders to risks associated with the future price of oil when there are contracts available to hedge the risks.
- 5.23** Microsoft is choosing an option on a portfolio of assets instead of the corresponding portfolio of options. The former is always less expensive because there is the potential for an increase in the price of one asset to be netted off against a decrease

in the price of another asset. Compare (a) an option with a strike price of \$20 on a portfolio of two assets each worth \$10 and (b) a portfolio of two options with a strike price of \$10, one on each of the assets. If both assets increase in price or both assets decrease in price, the payoffs are the same. But if one decreases and the other increases, the payoff from (a) is less than that from (b). Both the Asian feature and the basket feature in Microsoft's options help to reduce the cost of the options because of the possibility of gains and losses being netted.

- 5.24** It means that the price of the energy source can go up or down but will tend over time to get pulled back to its long-run average level. Electricity has the highest rate of mean reversion; oil has the lowest.
- 5.25** As we increase the frequency with which the asset price is observed, the asset price becomes more likely to hit the barrier and the value of a knock-out call goes down.
- 5.26** The average of the highest and lowest temperature each day is 75° Fahrenheit. The CDD each day is therefore 10 and the cumulative CDD for the month is $10 \times 31 = 310$. The payoff from the call option is therefore $(310 - 250) \times 5,000 = \$300,000$.
- 5.27** A 5×8 contract for May 2024 is a contract to provide electricity for five days per week during the off-peak period (11 P.M. to 7 A.M.). When daily exercise is specified, the holder of the option is able to choose each weekday whether he or she will buy electricity at the strike price. When there is monthly exercise, he or she chooses once at the beginning of the month whether electricity is to be bought at the strike price for the whole month. The option with daily exercise is worth more.
- 5.28** Two calculations are necessary to determine the initial margin. The first gives

$$500 \times (3.5 + 0.2 \times 57 - 3) = 5,950$$

The second gives

$$500 \times (3.5 + 0.1 \times 57) = 4,600$$

The initial margin is the greater of these, or \$5,950. Part of this can be provided by the initial amount of $\$500 \times 3.5 = \$1,750$ received for the options.

- 5.29** The cost of the shares is $\$500 \times 50$ or \$25,000. When shares are shorted, the proceeds of the sale form part of the margin. In this case, the total margin required is $1.6 \times \$25,000$, or \$40,000. The extra margin required is therefore \$15,000. This can be in the form of cash or marginable securities. When the share price rises to S , the value of the underlying position is $500S$. There is a margin call when

$$40,000 < 1.3 \times 500S \text{ or } S > 61.54.$$

- 5.30** A broker sets an initial margin and a maintenance margin. When the balance in the margin account falls below the maintenance margin level, the broker's client is required to bring it back up to the initial margin level. The exchange clearing house, when dealing with its members, sets the initial margin and maintenance margin to be the same. Every day the balance in the margin account must be at this level or more.

Chapter 6

- 6.1** When an ISDA master agreement is used, each side to an OTC derivatives transaction is taking the risk that the other side will default. When a CCP is used, the CCP acts as an intermediary so that each side has a transaction with the CCP.
- 6.2** Standard transactions between financial institutions must be cleared through CCPs. For transactions between financial institutions that are cleared bilaterally, both initial margin and variation margin are required. The initial margin must cover, with 99% confidence, market moves during a 10-day period of stressed market conditions.
- 6.3** The amount of collateral (typically cash or Treasury securities) that has to be provided by financial institutions for their derivatives books will increase.
- 6.4** The haircut applied to the market price of a financial instrument is the percentage reduction in its price that is made to determine its value for collateral purposes. For example, if an instrument is worth \$100 and is subject to a 10% haircut, it can be used to satisfy a \$90 collateral requirement.
- 6.5** An event of default occurs when one side to a derivatives transaction that is subject to an ISDA master agreement fails to post collateral or make a payment as required. An early termination follows an event of default and occurs when the nondefaulting party terminates all outstanding transactions with the defaulting party.
- 6.6** The exposure of A including its exposure to the CCP is reduced to 70. The exposure of A excluding its exposure to the CCP remains 0. The exposure of B including its exposure to the CCP is reduced to 100. The exposure of B excluding its exposure to the CCP is reduced to 70. The exposures of C are unaffected. The average exposure of the three parties, including the exposures to the CCP, is reduced from 110 to 86.7. The average exposure, excluding the exposures to the CCP, is reduced from 70 to 53.3.
- 6.7** The company will lose money if (a) the CCP defaults or (b) one of the other members defaults and not enough margin and default fund contribution has been provided by that member to cover losses when its positions are closed out.
- 6.8** Collateral requirements are normally calculated from the net mark-to-market value of all outstanding transactions. In determining settlement amounts, transactions with a positive value are netted against those with a negative value.
- 6.9** Rehypothecation occurs when collateral posted by A with B is used by B to meet collateral demands from C.
- 6.10** The activities of CCPs are much simpler than those of banks.
- 6.11** The mid-market values of transactions to the non-defaulting party at the time of the early termination adjusted upward by half the bid-ask spreads.

Chapter 7

- 7.1 Mortgages were frequently securitized. The only information that was retained during the securitization process was the applicant's FICO score and the loan-to-value ratio of the mortgage. The originators knew this and so this was the only information that they cared about.
- 7.2 There was a short-term imbalance between supply and demand because many people were persuaded to take out mortgages they could not afford.
- 7.3 When the loss rate on the mortgages is 5%, there are no losses on the mezzanine tranche of the ABS and therefore there are no losses on any of the tranches of the ABS CDO. When the loss rate on the mortgages is 12%, the loss on the mezzanine tranche of the ABS is $7/20$ or 35%. The loss rate on the equity and mezzanine tranches of the ABS CDO is 100%. The loss rate on the senior tranche of the ABS CDO is $10/75 = 13.33\%$.
- 7.4 Often a tranche is thin and the probability distribution of the loss is quite different from that on the bond. If a loss occurs, there is a high probability that it will be 100%. A 100% loss is much less likely for a bond.
- 7.5 An ABS is a set of tranches created from a portfolio of loans, bonds, credit card receivables, and so on. An ABS CDO is an ABS created from particular tranches (e.g., the BBB-rated tranches) of a number of different ABSs.
- 7.6 Investors underestimated how high the default correlations between mortgages would be in stressed market conditions. Investors also did not always realize that the tranches underlying ABS CDOs were usually quite thin so that they were either totally wiped out or untouched. There was an unfortunate tendency to assume that a tranche with a particular rating could be considered to be the same as a bond with that rating. This assumption was not valid for the reasons just mentioned.
- 7.7 *Agency costs* is a term used to describe the costs in a situation where the interests of two parties are not perfectly aligned. The incentives of traders, originators, valuers, the creators of the structured products, and the rating agencies arguably created agency costs.
- 7.8 The waterfall defines how the interest and principal cash flows from the underlying portfolio are distributed to the tranches. In a typical arrangement, interest cash flows are first used to pay the most senior tranche its promised return on its outstanding principal. The cash flows (if any) that are left over are used to provide the next most senior tranche with its promised return on its outstanding principal, and so on. Principal cash flows are usually used first to repay the most senior tranche, then the next most senior tranche, and so on. The equity tranche receives principal and interest payments only when more senior tranches have been paid.
- 7.9 An ABS CDO is created from tranches of ABSs (e.g., BBB tranches). A motivation on the part of the originator was that it could be difficult to find investors for the tranches directly.

- 7.10** Mian and Sufi showed that regions where mortgage application denials were highest in the United States in 1996 were also regions where mortgage origination grew particularly fast during the 2000 to 2006 period.
- 7.11** The mezzanine tranche of an ABS or ABS CDO is a tranche that is in the middle as far as seniority goes. It ranks below the senior tranches and therefore absorbs losses before they do. It ranks above the equity tranche (so that the equity tranche absorbs losses before it does).
- 7.12** As default correlation increases, the senior tranche of an ABS becomes more risky because it is more likely to suffer losses. The equity tranche becomes less risky as the default correlation increases. To understand this, note that in the limit when there is perfect correlation (and assuming no recoveries), all tranches have the same loss experience because either (a) all companies default or (b) no companies default. As we move toward the perfect correlation situation, the tranches become more similar so that the senior tranche becomes more risky and the equity tranche becomes less risky.
- 7.13** The end-of-year bonus reflects performance during the year. Traders and other employees of banks are liable to be always focusing on their next end-of-year bonus (and not on possible later effects of their decisions) and therefore have a short-term time horizon for their decision making. A trend toward deferred bonuses that can be clawed back is changing this to some extent.

Chapter 8

- 8.1** $2 \times \sqrt{3}$, or 3.46%.
- 8.2** The standard deviation of the percentage price change in one day is $25/\sqrt{252}$ or 1.57%, and 95% confidence limits are from -3.09% to +3.09%.
- 8.3** Volatility is much higher when markets are open than when they are closed. Traders therefore measure time in trading days rather than calendar days.
- 8.4** Implied volatility is the volatility that leads to the option price equaling the market price when Black–Scholes–Merton assumptions are used. It is found by trial and error. Because different options have different implied volatilities, traders are not using the same assumptions as Black–Scholes–Merton.
- 8.5** The standard formula for calculating standard deviation gives 0.547% per day. The simplified approach in equation (8.4) gives 0.530% per day.
- 8.6** The power law gives $0.01 = K \times 500^{-2}$ so that $K = 2,500$. (a) $2,500 \times 1,000^{-2} = 0.0025$ or 0.25%, and (b) $2,500 \times 2,000^{-2} = 0.000625$ or 0.0625%.
- 8.7** The variance rate estimate calculated at the end of day $n - 1$ for day n equals λ times the variance rate calculated at the end of day $n - 2$ for day $n - 1$ plus $1 - \lambda$ times the squared return on day $n - 1$. This is equivalent to assigning weights to squared returns that decline exponentially as we move back through time.

8.8 GARCH(1,1) adapts the EWMA model by giving some weight to a long-run average variance rate. Whereas the EWMA has no mean reversion, GARCH(1,1) is consistent with a mean-reverting variance rate model.

8.9 In this case, $\sigma_{n-1} = 0.015$ and $u_{n-1} = 0.5/30 = 0.01667$, so that equation (10.8) gives

$$\sigma_n^2 = 0.94 \times 0.015^2 + 0.06 \times 0.01667^2 = 0.0002281$$

The volatility estimate on day n is therefore $\sqrt{0.0002281} = 0.015103$ or 1.5103%.

8.10 Reducing λ from 0.95 to 0.85 means that more weight is given to recent observations of u_i^2 and less weight is given to older observations. Volatilities calculated with $\lambda = 0.85$ will react more quickly to new information and will bounce around much more than volatilities calculated with $\lambda = 0.95$.

8.11 With the usual notation $u_{n-1} = 20/1,040 = 0.01923$ so that

$$\sigma_n^2 = 0.000002 + 0.06 \times 0.01923^2 + 0.92 \times 0.01^2 = 0.0001162$$

This gives $\sigma_n = 0.01078$. The new volatility estimate is therefore 1.078% per day.

8.12 The proportional daily change is $-0.005/1.5000 = -0.003333$. The current daily variance estimate is $0.006^2 = 0.000036$. The new daily variance estimate is

$$0.9 \times 0.000036 + 0.1 \times 0.003333^2 = 0.000033511$$

The new volatility is the square root of this. It is 0.00579 or 0.579%.

8.13 The weight given to the long-run average variance rate is $1 - \alpha - \beta$ and the long-run average variance rate is $\omega/(1 - \alpha - \beta)$. Increasing ω increases the long-run average variance rate. Increasing α increases the weight given to the most recent data item, reduces the weight given to the long-run average variance rate, and increases the level of the long-run average variance rate. Increasing β increases the weight given to the previous variance estimate, reduces the weight given to the long-run average variance rate, and increases the level of the long-run average variance rate.

8.14 The long-run average variance rate is $\omega/(1 - \alpha - \beta)$ or $0.000004/0.03 = 0.0001333$. The long-run average volatility is $\sqrt{0.0001333}$ or 1.155%. The equation describing the way the variance rate reverts to its long-run average is

$$E[\sigma_{n+k}^2] = V_L + (\alpha + \beta)^k(\sigma_n^2 - V_L)$$

In this case,

$$E[\sigma_{n+k}^2] = 0.0001333 + 0.97^k(\sigma_n^2 - 0.0001333)$$

If the current volatility is 20% per year, $\sigma_n = 0.2/\sqrt{252} = 0.0126$. The expected variance rate in 20 days is

$$0.0001333 + 0.97^{20}(0.0126^2 - 0.0001333) = 0.0001471$$

The expected volatility in 20 days is therefore $\sqrt{0.0001471} = 0.0121$ or 1.21% per day.

- 8.15** The FTSE expressed in dollars is XY where X is the FTSE expressed in sterling and Y is the exchange rate (value of one pound in dollars). Define x_i as the proportional change in X on day i and y_i as the proportional change in Y on day i . The proportional change in XY is approximately $x_i + y_i$. The standard deviation of x_i is 0.018 and the standard deviation of y_i is 0.009. The correlation between the two is 0.4. The variance of $x_i + y_i$ is therefore

$$0.018^2 + 0.009^2 + 2 \times 0.018 \times 0.009 \times 0.4 = 0.0005346$$

so that the volatility of $x_i + y_i$ is 0.0231 or 2.31%. This is the volatility of the FTSE expressed in dollars. Note that it is greater than the volatility of the FTSE expressed in sterling. This is the impact of the positive correlation. When the FTSE increases, the value of sterling measured in dollars also tends to increase. This creates an even bigger increase in the value of FTSE measured in dollars. Similarly, when FTSE decreases, the value of sterling measured in dollars also tends to decrease, creating an even bigger decrease in the value of FTSE measured in dollars.

- 8.16** In this case, $V_L = 0.000003/0.02 = 0.00015$ and equation (10.14) gives the expected variance rate in 30 days as

$$0.00015 + 0.98^{30}(0.01^2 - 0.00015) = 0.000123$$

The volatility is $\sqrt{0.000123} = 0.0111$ or 1.11% per day.

- 8.17** In this case, $V_L = 0.000002/0.02 = 0.0001$. In equation (10.15), $V_L = 0.0001$, $a = 0.0202$, $T = 20$, and $V(0) = 0.000169$ so that the volatility is 19.88%.

Chapter 9

- 9.1** You need the standard deviations of the two variables.
- 9.2** Loosely speaking, correlation measures the extent of linear dependence. It does not measure other types of dependence. When $y = x^2$, there is clearly dependence between x and y . However, $E(xy) = E(x^3)$. Both $E(x)$ and $E(x^3)$ are zero when x is normal (or when it has any symmetrical distribution centered at zero). It follows that the coefficient of correlation between x and y is zero.
- 9.3** In a factor model, the correlation between two variables arises entirely because of their correlation with other variables known as factors. A factor model reduces the number of estimates that have to be made when correlations between large numbers of variables are being produced.
- 9.4** A positive-semidefinite matrix is a matrix that satisfies equation (9.4) for all vectors $\mathbf{w}$. If a correlation matrix is not positive-semidefinite, the correlations are internally inconsistent.
- 9.5 (a)** The volatilities and correlation imply that the current estimate of the covariance is $0.25 \times 0.016 \times 0.025 = 0.0001$.

- (b) If the prices of the assets at close of trading are \$20.50 and \$40.50, the proportional changes are $0.5/20 = 0.025$ and $0.5/40 = 0.0125$. The new covariance estimate is

$$0.95 \times 0.0001 + 0.05 \times 0.025 \times 0.0125 = 0.0001106$$

The new variance estimate for asset A is

$$0.95 \times 0.016^2 + 0.05 \times 0.025^2 = 0.00027445$$

so that the new volatility is 0.0166. The new variance estimate for asset B is

$$0.95 \times 0.025^2 + 0.05 \times 0.0125^2 = 0.000601562$$

so that the new volatility is 0.0245. The new correlation estimate is

$$\frac{0.0001106}{0.0166 \times 0.0245} = 0.272$$

- 9.6 The most recent returns for X and Y are $1/30 = 0.03333$ and $1/50 = 0.02$, respectively. The previous covariance is $0.01 \times 0.012 \times 0.50 = 0.00006$. The new estimate of the covariance is

$$0.000001 + 0.04 \times 0.03333 \times 0.02 + 0.94 \times 0.00006 = 0.0000841$$

The new estimate of the variance of X is

$$0.000003 + 0.04 \times 0.03333^2 + 0.94 \times 0.01^2 = 0.0001414$$

so that the new volatility of X is $\sqrt{0.0001414} = 0.01189$ or 1.189%. The new estimate of the variance of Y is

$$0.000003 + 0.04 \times 0.02^2 + 0.94 \times 0.012^2 = 0.0001544$$

so that the new volatility of Y is $\sqrt{0.0001544} = 0.01242$ or 1.242%. The new estimate of the correlation between the assets is therefore $0.0000841 / (0.01189 \times 0.01242) = 0.569$.

- 9.7 Continuing with the notation in the answer to Problem 8.15, define z_i as the proportional change in the value of the S&P 500 on day i . The covariance between x_i and z_i is

$$0.7 \times 0.018 \times 0.016 = 0.0002016$$

The covariance between y_i and z_i is

$$0.3 \times 0.009 \times 0.016 = 0.0000432$$

The covariance between $x_i + y_i$ and z_i equals the covariance between x_i and z_i plus the covariance between y_i and z_i . It is

$$0.0002016 + 0.0000432 = 0.0002448$$

The volatility of $x_i + y_i$ is 2.31% from Problem 8.15. The correlation between $x_i + y_i$ and z_i is

$$\frac{0.0002448}{0.016 \times 0.0231} = 0.662$$

Note that the volatility of the S&P 500 drops out in this calculation.

9.8

V_1	V_2		
	0.25	0.5	0.75
0.25	0.095	0.163	0.216
0.50	0.163	0.298	0.413
0.75	0.216	0.413	0.595

9.9 The formulas are

$$\epsilon_1 = z_1, \quad \epsilon_2 = \rho_{12}z_1 + z_2\sqrt{1 - \rho_{12}^2}, \quad \epsilon_3 = \alpha_1z_1 + \alpha_2z_2 + \alpha_3z_3$$

where

$$\alpha_1 = \rho_{13}, \quad \alpha_1\rho_{12} + \alpha_2\sqrt{1 - \rho_{12}^2} = \rho_{23}, \quad \alpha_1^2 + \alpha_2^2 + \alpha_3^2 = 1$$

This means that

$$\alpha_1 = \rho_{13}, \quad \alpha_2 = \frac{\rho_{23} - \rho_{13}\rho_{12}}{\sqrt{1 - \rho_{12}^2}}, \quad \alpha_3 = \sqrt{1 - \alpha_1^2 - \alpha_2^2}$$

9.10 Tail dependence is the tendency for extreme values for two or more variables to occur together. The choice of the copula affects tail dependence. For example, the Student's t -copula gives more tail dependence than the Gaussian copula.

9.11 Sample from a bivariate Student's t distribution as in Section 9.5.2. Convert each sample to a normal distribution on a percentile-to-percentile basis.

9.12 The probability that $V_1 < 0.1$ is 0.05. The conditional probability that $V_2 < 0.1$ is $0.006/0.05 = 0.12$. The conditional probability that $V_2 < 0.2$ is $0.017/0.05 = 0.34$, and so on.

9.13 When $V_1 = 0.2$, $U_1 = -0.8416$. From the properties of the bivariate normal distribution, the median of U_2 is $-0.5 \times 0.8416 = -0.4208$. The percentile of the conditional median is therefore $N(-0.4208) = 0.3369$. This translates into a median value for V_2 of x where $2x^2 = 0.3369$. The conditional median value for V_2 is therefore 0.4105.

9.14 In this case,

$$\text{WCDR}(T, X) = N\left(\frac{N^{-1}(0.015) + \sqrt{0.2}N^{-1}(0.995)}{\sqrt{1 - 0.2}}\right) = 0.127$$

The loss rate that we are 99.5% certain will not be exceeded is 12.7%.

9.15 The maximum likelihood estimates for the probability of default and the copula correlation are 3.92% and 11.18%, respectively.

Chapter 10

- 10.1** From equation (10.3), the probability is $N(d_2)$ where

$$d_2 = \frac{\ln(50/70) + (0.12 - 0.2^2/2) \times 2}{0.2\sqrt{2}} = -0.4825$$

The probability is therefore $N(-0.4825) = 0.315$.

- 10.2** From equation (10.4), this is

$$50 \times \exp[(0.12 - 0.2^2/2) \times 2 - N^{-1}(0.05) \times 0.2 \times \sqrt{2}] = 97.25$$

or \$97.25.

- 10.3** The risk-neutral valuation principle states that, if we value a derivative on the assumption that investors are risk-neutral (i.e., investors do not require a higher expected return for bearing risks), we get the right answer for all worlds (including the real world).
- 10.4** The expected value is higher in the real world. The higher value reflects the compensation for risk required by the investor.
- 10.5** By buying both derivatives, one is certain to earn \$2 on an initial investment of \$98. The risk-free rate is therefore $2/98 = 2.04\%$ with annual compounding. The risk-neutral probability of default is π where $100\pi/(1 + 0.0204) = 3$ or $\pi = 3.0612\%$.
- 10.6** The risk-neutral probability of the payoff is given by equation (10.3) with $V = 30$, $S_0 = 25$, $\sigma = 0.3$, $\mu = 0.03$, and $T = 0.25$. It is

$$N \left[\frac{(25/30) + (0.03 - 0.3^2/2) \times 0.25}{0.3 \times \sqrt{0.25}} \right] = 0.1074$$

The value of the option is therefore $100 \times 0.1074 \times e^{-0.03 \times 0.25} = \10.66 . For the real-world probability of a payoff, we set $\mu = 0.1$ instead of 0.03 in the equation to get 0.1305.

- 10.7** The real world is used to simulate possible values of the underlying variables in one year. The risk-neutral world is then used to calculate the value of the portfolio at the one-year point for each simulation trial.
- 10.8** *Mean reversion* is the term used to describe the tendency for the values of variables such as interest rates and commodity prices to be pulled back to a central value. Of course, an uncertain random movement is superimposed on this pull-back.
- 10.9** Girsanov's theorem tells us that when we move from the real world to the risk-neutral world or vice versa, the expected growth rates of market variables change but their volatilities remain the same.

Chapter 11

- 11.1** VaR is the loss that is not expected to be exceeded with a certain confidence level. Expected shortfall is the expected loss conditional that the loss is worse than the VaR level. Expected shortfall has the advantage that it always satisfies the subadditivity (diversification is good) condition.
- 11.2** A spectral risk measure is a risk measure that assigns weights to the quantiles of the loss distribution. For the subadditivity condition to be satisfied, the weight assigned to the q th quantile must be a non-decreasing function of q .
- 11.3** There is a 5% chance that you will lose \$6,000 or more during a one-month period.
- 11.4** Your expected loss during a bad month is \$6,000. Bad months are defined as the months where returns are less than the five-percentile point on the distribution of monthly returns.
- 11.5** (a) \$1 million.
 (b) The expected shortfall is $0.9 \times 10 + 0.1 \times 1$ or \$9.1 million.
 (c) There is a probability of $0.009^2 = 0.000081$ of a loss of \$20 million, a probability of $2 \times 0.009 \times 0.991 = 0.017838$ of a loss of \$11 million, and a probability of $0.991^2 = 0.982081$ of a loss of \$2 million. The VaR when the confidence level is 99% is therefore \$11 million.
 (d) The expected shortfall is $(0.000081 \times 20 + 0.009919 \times 11)/0.01 = \11.07 million.
 (e) Because $1 + 1 < 11$, the subadditivity condition is not satisfied for VaR. Because $9.1 + 9.1 > 11.07$, it is satisfied for expected shortfall.
- 11.6** (a) $2 \times 1.96 = \$3.92$ million, (b) $\sqrt{5} \times 2 \times 1.96 = \8.77 million, (c) $\sqrt{5} \times 2 \times 2.33 = 10.40$ million.
- 11.7** (b) becomes \$9.96 million and (c) becomes \$11.82 million.
- 11.8** Marginal VaR is the rate of change of VaR with the amount invested in the i th asset. Incremental VaR is the incremental effect of the i th asset on VaR (i.e., the difference between VaR with and without the asset). Component VaR is the part of VaR that can be attributed the i th asset (the sum of component VaRs equals the total VaR).
- 11.9** The probability of 17 or more exceptions is $1 - \text{BINOMDIST}(16, 1000, 0.01, \text{TRUE})$ or 2.64%. The model should be rejected at the 5% confidence level.
- 11.10** Bunching is the tendency for exceptions to be clustered together rather than occurring randomly throughout the time period considered.
- 11.11** We are interested in the variance of $\Delta P_1 + \Delta P_2 + \dots + \Delta P_T$. This is $\sum_{i=1}^T \sigma_i^2 + 2 \sum_{i>j} \rho_{ij} \sigma_i \sigma_j$ where σ_i is the standard deviation of ΔP_i and ρ_{ij} is the correlation between ΔP_i and ΔP_j . In this case, $\sigma_i = \sigma$ for all i and $\rho_{ij} = \rho^{i-j}$ when $i > j$. After further algebraic manipulations, this leads directly to equation (11.5).
- 11.12** The standard deviation in three months is $2\sqrt{3} = \$3.464$ million. Also, $N^{-1}(0.98) = 2.054$. The 98% three-month VaR is therefore $3.464 \times 2.054 =$

\$7.11 million. The 98% three-month ES is $3.464 \times \exp(-2.054^2/2)/(\sqrt{2\pi} \times 0.02) = \8.39 million.

Chapter 12

12.1 The assumption is that the statistical process driving changes in market variables over the next day is the same as that over the past 500 days.

12.2

$$\lambda^{i-1}(1-\lambda)/(1-\lambda^n) = \frac{\lambda^{i-1}}{1+\lambda+\lambda^2+\dots+\lambda^{n-1}}$$

This shows that as λ approaches 1, the weights approach $1/n$.

12.3 The standard error of the estimate is

$$\frac{1}{0.01} \sqrt{\frac{0.05 \times 0.95}{1,000}} = 0.69$$

The standard error of the VaR estimate is \$0.69 million.

12.4 (a) The 95% one-day VaR is the 25th worst loss. This is \$163,620. (b) The 95% one-day ES is the average of the 24 highest losses. It is \$320,360. (c) The 97% one-day VaR is the 15th worst loss. This is \$229,683. (d) The 97% one-day ES is the average of the 14 highest losses. It is \$428,667.

12.5 In worksheet 2 (Scenarios), the portfolio investments are changed to 2,500 in cells L2:O2. The losses are then sorted from the largest to the smallest. The fifth worst loss is \$394,437. This is the one-day 99% VaR. The average of the four worst losses is \$693,536. This is the one-day 99% ES.

12.6 The value of λ in cell F2 of worksheet 4 (Scenarios with Weights) is changed from 0.995 to 0.99. The losses are then sorted from the largest to the smallest. The weights for the worst three scenarios are 0.004833, 0.004931, and 0.004690. The cumulative weights are 0.004833, 0.009764, and 0.014454. The VaR is therefore the third-highest loss or \$653,541. The ES is $0.4833 \times 922,484 + 0.4931 \times 858,423 + (1 - 0.4833 - 0.4931) \times 653,541$ or \$884,560.

12.7 The value of λ in cell K1 of worksheet 9 (Losses Adjusted for Loss SD) is changed from 0.94 to 0.96. The adjusted losses are then sorted from the highest to the lowest. The fifth worst loss is \$596,180. This is the new one-day 99% VaR. The average of the four highest losses is \$812,948. This is the new one-day 99% ES.

12.8 This is

$$\frac{25}{500} \left[1 + 0.354 \frac{400 - 160}{110.46} \right]^{-1/0.354}$$

or 0.00998.

12.9 The VaR (\$000s) is

$$160 + \frac{110.46}{0.354} \left\{ \left[\frac{500}{25} (1 - 0.97) \right]^{-0.354} - 1 \right\} = 221.85$$

or \$221,850.

- 12.10** The maximum likelihood estimates of ξ and β become 0.391 and 100.13. The one-day 99% VaR becomes \$396,210 and the one-day 99.9% VaR becomes \$1,129,996. The one-day 99% ES becomes \$718,825 and the one-day 99.9% ES becomes \$1,923,975.
- 12.11** The ranked losses from the volatility scaling procedure must be transferred to worksheet 11 (Extreme Val Theory). With $u = 250$, we have $n_u = 28$ and the maximum likelihood values of ξ and β are 0.077 and 139.14. The one-day 99% VaR is \$506,314. The one-day 99.9% VaR is \$906,461. The one-day ES with 99% confidence is \$678,397. The one-day ES with 99.9% confidence is \$1,111,872. The probability of a loss greater than \$600,000 is

$$\frac{28}{500} \left[1 + 0.077 \frac{600 - 250}{139.14} \right]^{-1/0.077} = 0.0056$$

Chapter 13

- 13.1** The standard deviation of the daily change in the investment in each asset is \$1,000. The variance of the portfolio's daily change is

$$1,000^2 + 1,000^2 + 2 \times 0.3 \times 1,000 \times 1,000 = 2,600,000$$

The standard deviation of the portfolio's daily change is the square root of this, or \$1,612.45. The five-day 97% value at risk is therefore $1.8808 \times \sqrt{5} \times 1,612.45 = \$6,781$. The five-day 97% ES is

$$\sqrt{5} \times 1,612.45 \times \frac{e^{-1.8808^2/2}}{\sqrt{2\pi} \times 0.03} = \$8,178$$

- 13.2** $\Delta P = 3.9\Delta x$. The standard deviation of ΔP is $3.9 \times 0.007 = 0.027$. It follows that the 10-day 99% VaR for the portfolio is $0.027 \times 2.326 \times \sqrt{10} = 0.20$.
- 13.3** The relationship is $\Delta P = 3.9\Delta x + 0.5 \times 4.3 \times \Delta x^2$ or $\Delta P = 3.9\Delta x + 2.15\Delta x^2$.
- 13.4** Suppose that the risk measure (VaR or ES) for a portfolio is its standard deviation multiplied by β . The risk weight of a risk factor is its standard deviation multiplied by β . The risk sensitivity for a portfolio is the amount by which the portfolio changes for a 1% or one basis point change in the risk factor.
- 13.5** The daily variance of the portfolio is

$$6^2 \times 20^2 + 4^2 \times 8^2 = 15,424$$

and the daily standard deviation is $\sqrt{15,424} = \$124.19$. Since $N^{-1}(0.9) = 1.282$, the five-day 90% value at risk is

$$124.19 \times \sqrt{5} \times 1.282 = \$355.89$$

- 13.6** The change in the value of an option is not linearly related to the changes in the values of the underlying variables. When the changes in the values of underlying variables are normal, the change in the value of the option is non-normal. The linear model assumes that it is normal and is, therefore, only an approximation.
- 13.7** The contract is a long position in a sterling bond combined with a short position in a dollar bond. The value of the sterling bond is $1.53e^{-0.05 \times 0.5}$, or \$1.492 million. The value of the dollar bond is $1.5e^{-0.05 \times 0.5}$, or \$1.463 million. The variance of the change in the value of the contract in one day is

$$1.492^2 \times 0.0006^2 + 1.463^2 \times 0.0005^2 - 2 \times 0.8 \times 1.492 \times 0.0006 \times 1.463 \times 0.0005 = 0.000000288$$

The standard deviation is therefore \$0.000537 million. The 10-day 99% VaR is $0.000537 \times \sqrt{10} \times 2.33 = \0.00396 million.

- 13.8** The alphas should be changed to 2,500. This changes the one-day 99% VaR to \$275,757 and the one-day ES to \$315,926 when volatilities and correlations are estimated using the equally weighted model. It changes the one-day 99% VaR to \$313,219 and the one-day 99% ES to \$358,844 when EWMA with $\lambda = 0.94$ is used.
- 13.9** This changes the one-day 99% VaR from \$302,459 to \$389,300 and the one-day ES from \$346,516 to \$450,590.

Chapter 14

- 14.1** In this case, the interest rate mismatch is \$10 billion. The bank's net interest income declines by \$100 million each year for the next three years.
- 14.2** If long-term rates were simply a reflection of expected future short-term rates, we would expect long rates to be less than short rates as often as they are greater than short rates. (This is based on the assumption that half of the time investors expect rates to increase, and half of the time investors expect rates to decrease.) Liquidity preference theory argues that long-term rates are high relative to expected future short-term rates. This means that long rates are greater than short rates most of the time. When long rates are less than short rates, the market is expecting a relatively steep decline in rates.
- 14.3** The two alternative procedures mentioned in the chapter for handling interest rates when the model-building approach is used to calculate VaR involve (a) the use of principal components analysis and (b) describing the term structure with 10 vertices.
- 14.4** The 10-year vertex has a sensitivity of \$30,000 because a one-basis-point increase in that rate leads to an increase of 0.6 basis points in the 12-year rate. Similarly, the 15-year vertex has a sensitivity of \$20,000. The other vertices have no sensitivity.

- 14.5** The variance of the change in the value of the portfolio in one day is (in \$000s)

$$(10 \times 7)^2 + (8 \times 8)^2 + 2 \times 10 \times 7 \times 8 \times 8 \times 0.8 = 16,164$$

Because $N^{-1}(0.98) = 2.054$, the ES is therefore

$$\frac{\sqrt{16,164} \times 5e^{-2.054^2/2}}{\sqrt{2\pi} \times 0.02} = 688.2$$

or \$688,200.

- 14.6** Duration provides information about the effect of a small parallel shift in the yield curve on the value of a bond portfolio. The percentage decrease in the value of the portfolio equals the duration of the portfolio multiplied by the amount by which interest rates are increased in the small parallel shift. Its main limitation is that it applies only to parallel shifts in the yield curve that are small.

- 14.7** (a) The bond's price is 86.80, (b) the bond's duration is 4.256 years, (c) the duration formula shows that when the yield decreases by 0.2% the bond's price increases by 0.74, and (d) recomputing the bond's price with a yield of 10.8% gives a price of 87.54, which is approximately consistent with (a) and (c).

- 14.8** (a) The bond's price is 88.91, (b) the bond's modified duration is 3.843 years, (c) the duration formula shows that when the yield decreases by 0.2% the bond's price increases by 0.68, and (d) recomputing the bond's price with a yield of 10.8% (annually compounded) gives a price of 89.60, which is approximately consistent with (a) and (c).

- 14.9** The bond's price is 104.80. The duration of the bond is 5.35. The convexity is 30.60. The effect of a 1% increase in the yield is estimated by equation (9.4) as

$$104.80 \times (-0.01 \times 5.35 + 0.5 \times 30.60 \times 0.0001) = -5.44$$

The bond's price actually changes to 99.36, which is consistent with the estimate.

- 14.10** We can (a) perturb points on the yield curve (see Figure 14.6), (b) perturb sections of the yield curve (see Figure 14.8), and (c) perturb the market quotes used to create the yield curve.

- 14.11** Partial durations enable the impact of any non-parallel shift in the yield curve to be evaluated. Durations are concerned with parallel shifts.

- 14.12** The impact on the portfolio, measured as a proportion of the value of the portfolio, is

$$\begin{aligned} &-(0.2 \times 0.001 + 0.6 \times 0.0008 + 0.9 \times 0.0007 + 1.6 \times 0.0006 \\ &+ 2.0 \times 0.0005 - 2.1 \times 0.0003 - 3.0 \times 0.0001) = -0.00234 \end{aligned}$$

The portfolio decreases by 0.234%.

- 14.13** Dollar duration is defined as the product of the duration of a portfolio and its value. Dollar convexity is defined as the product of the convexity of a portfolio and its value.

- 14.14** The partial durations add up to the total duration. The DV01 is the total duration times the portfolio value times 0.0001.

Chapter 15

- 15.1** The value of the portfolio decreases by \$10,500.
- 15.2** The value of the portfolio increases by \$400.
- 15.3** In both cases, it increases by $0.5 \times 30 \times 2^2$ or \$60.
- 15.4** A delta of 0.7 means that, when the price of the stock increases by a small amount, the price of the option increases by 70% of this amount. Similarly, when the price of the stock decreases by a small amount, the price of the option decreases by 70% of this amount. A short position in 1,000 options has a delta of -700 and can be made delta neutral with the purchase of 700 shares.
- 15.5** A theta of -100 per day means that if one day passes with no change in either the stock price or its volatility, the value of the option position declines by \$100. If a trader feels that neither the stock price nor its implied volatility will change, she should write an option with as big a negative theta as possible so that her position has a high positive theta. Relatively short-life, at-the-money options have the biggest negative theta.
- 15.6** The gamma of an option position is the rate of change of the delta of the position with respect to the asset price. For example, a gamma of 0.1 would indicate that, when the asset price increases by a certain small amount, delta increases by 0.1 of this amount. When the gamma of an option writer's position is large and negative and the delta is zero, the option writer will lose significant amounts of money if there is a large movement (either an increase or a decrease) in the asset price.
- 15.7** To hedge an option position, it is necessary to create the opposite option position synthetically. For example, to hedge a long position in a put, it is necessary to create a short position in a put synthetically. It follows that the procedure for creating an option position synthetically is the reverse of the procedure for hedging the option position.
- 15.8** A long position in either a put or a call option has a positive gamma. From Figure 15.9, when gamma is positive, the hedger gains from a large change in the stock price and loses from a small change in the stock price. Hence the hedger will fare better in case (b). When the portfolio contains short option positions, the hedger will fare better in case (a).
- 15.9** The delta indicates that when the value of the euro exchange rate increases by \$0.01, the value of the bank's position increases by $0.01 \times 30,000 = \$300$. The gamma indicates that when the euro exchange rate increases by \$0.01, the delta of the portfolio decreases by $0.01 \times 80,000 = 800$. For delta neutrality, 30,000 euros should be shorted. When the exchange rate moves up to 0.93, we expect the delta of the portfolio to decrease by $(0.93 - 0.90) \times 80,000 = 2,400$ so that it becomes 27,600. To maintain delta neutrality, it is therefore necessary for the bank to unwind its short position by 2,400 euros so that a net 27,600 euros have been shorted. When a portfolio is delta neutral and has a negative gamma, a loss

is experienced when there is a large movement in the underlying asset price. We can conclude that the bank is likely to have lost money.

- 15.10** When used in the way described in the text, it does assume volatility is constant. In theory, we could implement a static options replication strategy where there are three dimensions: time, the stock price, and volatility. Prices are then matched on a surface in the three-dimensional space.
- 15.11** Ten regular options are likely to be needed. This is because there are 10 equations to be satisfied, one for each point on the boundary.
- 15.12** The payoff from an Asian option becomes more certain with the passage of time. As a result, the amount of uncertainty that needs to be hedged decreases with the passage of time.
- 15.13** Consider a portfolio of options dependent on a single market variable. A single trade is all that is necessary to make the position delta neutral regardless of the number of options in the portfolio.
- 15.14** The price, delta, gamma, vega, theta, and rho are 0.0217, -0.396 , 5.415, 0.00203, -0.0000625 , and -0.00119 . Delta predicts that the option price should decrease by approximately 0.000396 when the exchange rate increases by 0.001. This is what we find. When the exchange rate is increased to 0.751, the option price decreases to 0.0213.

Chapter 16

- 16.1** Scenarios can be generated by (a) making large changes to key variables such as interest rates, equity prices, and so on; (b) making percentage changes in all market variables that are the same as those that occurred on particular extreme days in the past; and (c) asking a committee of senior management to generate the scenarios.
- 16.2** Reverse stress testing involves using an algorithm to search for scenarios that lead to big losses. It is used to help identify appropriate scenarios for stress testing.
- 16.3** Financial institutions might consider that regulators will require more capital if scenarios leading to large losses are considered.
- 16.4** Traffic light options were options that provided payoffs when the scenarios considered by insurance company regulators occurred. The danger is that, when a financial institution buys traffic light options, it is protecting itself against a too narrow range of adverse scenarios. The financial institution might not be protected against a scenario that is similar to, but not exactly the same as, the scenarios covered by the traffic light options.
- 16.5** Senior managers are in the best position to develop scenarios for stress testing. Their involvement in the development of scenarios makes it more likely that they will take the results of stress testing seriously and incorporate stress testing in their strategic decision making.

- 16.6** The advantage is that the same scenarios are considered by different banks and systemic risks can be evaluated. The scenarios might be worse than those considered by the banks themselves (see Problem 16.3). The disadvantage is that it might lead to the banks themselves not spending as much time as they should on stress testing.
- 16.7** An objective probability is calculated from data. A subjective probability is a “degree of belief” and reflects a person’s judgment.
- 16.8** The total probability of the stressed scenarios is 1.5%. The probability associated with the historical simulation scenarios is therefore 98.5%. Each historical simulation scenario has a probability of 0.197. When the scenarios and their probabilities are ranked, we see that the VaR with a 99% confidence limit is \$653,541. (For a loss of \$700,000, the cumulative probability is 0.00943. For a loss of \$653,541, the cumulative probability is 0.01141.)
- 16.9** In this case, the position values are 941.34, −164.39, −1,349.94, and −78.36. The worst-case scenario is where the asset price is 60 and its volatility is 30%. This leads to a loss of \$341.39.

Chapter 17

- 17.1** The investment grade ratings are AAA, AA, A, BBB. AA, A, and BBB and are subdivided as explained in Section 1.7.
- 17.2** A hazard rate $h(t)$ at time t is such that $h(t)dt$ is the probability of default between times t and $t + dt$ conditional on no default prior to time t .
- 17.3** From equation (17.2), the average hazard rate, $\bar{\lambda}$, satisfies $0.0334 = 1 - e^{-\bar{\lambda} \times 1}$ so that $\bar{\lambda} = -\ln(0.9666) = 0.0340$. The average hazard rate is 3.40% per year.
- 17.4** Conditional on no default by year 2, the probability of a default in year 3 is

$$(0.0346 - 0.0193)/(1 - 0.0193) = 0.0156$$

Average hazard rate for the third year, $\bar{\lambda}$, satisfies $1 - e^{-\bar{\lambda} \times 1} = 0.0156$. It is 1.57% per year.

- 17.5** The seller receives $300,000,000 \times 0.0060 \times 0.5 = \$900,000$ at times 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, and 4.0 years. The seller also receives a final accrual payment of about \$300,000 ($= \$300,000,000 \times 0.0060 \times 2/12$) at the time of the default (four years and two months). The seller pays $300,000,000 \times 0.6 = \$180,000,000$ at the time of the default.
- 17.6** Sometimes there is physical settlement of a credit default swap, and sometimes there is cash settlement. In the event of a default when there is physical settlement, the buyer of protection sells bonds issued by the reference entity for their face value. Bonds with a total face value equal to the notional principal can be sold. In the event of a default when there is cash settlement, a calculation agent or an auction determines the value of the cheapest-to-deliver bonds issued by the reference entity a specified number of days after the default event. The cash payoff

is then based on the excess of the face value of these bonds over the estimated value.

- 17.7** Risk-neutral default probabilities are backed out from credit spreads. Real-world default probabilities are calculated from historical data. Risk-neutral default probabilities should be used for valuation. Real-world default probabilities should be used for scenario analysis and credit VaR calculations.
- 17.8** The payoff is $L(1 - R)$ where L is the notional principal and R is the recovery rate.
- 17.9** From equation (17.3), the average hazard rate over the three years is $0.0050/(1 - 0.3) = 0.0071$ or 0.71% per year.
- 17.10** From equation (17.3), the average hazard rate over five years is $0.0080/(1 - 0.4)$ or 1.333% per year. Similarly, the average hazard rate over three years is 1.1667% per year. This means that the average hazard rate for years 4 and 5 is $(5 \times 1.333 - 3 \times 1.1667)/2 = 1.58\%$.
- 17.11** Real-world probabilities of default should be used for calculating credit value at risk. Risk-neutral probabilities of default should be used for adjusting the price of a derivative for default.
- 17.12** The recovery rate for a bond is the value of the bond shortly after the issuer defaults as a percentage of its face value.
- 17.13** The first number in the second column of Table 17.5 is calculated as

$$-\frac{1}{7} \ln(1 - 0.0051) = 0.00073$$

or 0.073% per year. The second number in the second column of Table 17.5 is calculated as

$$-\frac{1}{7} \ln(1 - 0.0049) = 0.00070$$

and so on. The final number is

$$-\frac{1}{7} \ln(1 - 0.5075) = 0.10118$$

- 17.14** The no-default value of the bond

$$2e^{-0.03 \times 0.5} + 2e^{-0.03 \times 1.0} + \dots + 102e^{-0.03 \times 4.0} = 103.66$$

The market price is 96.16. An analysis similar to that in Table 17.3 shows that if Q is the default probability per year, the loss from defaults is $272.69Q$. The implied probability of default is therefore given by solving $103.66 - 96.16 = 272.69Q$. It is 2.74% per year.

- 17.15** If Q_1 is the default probability at times 0.5, 1.5, and 2.5 years, an analysis similar to that in Table 17.3 shows that the present value of the loss from defaults for the first bond is $178.31Q_1$ per \$100 of face value. The default-free value of the bond is

$$4 \times e^{-0.035 \times 1} + 4 \times e^{-0.035 \times 2} + 104 \times e^{-0.035 \times 3} = 101.23$$

and the market value is 98.35. It follows that $178.31Q_1 = 101.23 - 98.35$ so that $Q_1 = 0.0157$ or 1.57%. If Q_2 is the probability of default at times 3.5 and 4.5, an

analysis similar to that in Table 17.3 gives the present value of the loss from default for the second bond to be $180.56Q_1 + 108.53Q_2$. The default-free value of the second bond is

$$4e^{-0.035 \times 1} + 4 \times e^{-0.035 \times 2} + \dots + 104 \times e^{-0.035 \times 5} = 101.97$$

The market value is 96.24. It follows that $180.56Q_1 + 108.53Q_2 = 101.97 - 96.24$. Substituting for Q_1 , we find that $Q_2 = 0.0260$, or 2.60%.

- 17.16** We can assume that the principal is paid and received at the end of the life of the swap without changing the swap's value. If the spread were zero, the present value of the floating payments per dollar of principal would be 1. The payment of floating plus the spread therefore has a present value of $1 + V$. The payment of the bond cash flows has a present value per dollar of principal of B^* . The initial payment required from the payer of the bond cash flows per dollar of principal is $1 - B$. (This may be negative; an initial amount of $B - 1$ is then paid by the payer of the floating rate.) Because the asset swap is initially worth zero, we have

$$1 + V = B^* + 1 - B$$

so that

$$V = B^* - B$$

- 17.17** The value of the debt in Merton's model is $V_0 - E_0$ or

$$V_0 - V_0 N(d_1) + De^{-rT} N(d_2) = De^{-rT} N(d_2) + V_0 N(-d_1)$$

If the credit spread is s , this should equal $De^{-(r+s)T}$ so that

$$De^{-(r+s)T} = De^{-rT} N(d_2) + V_0 N(-d_1)$$

Substituting $De^{-rT} = LV_0$, we get

$$Le^{-sT} = LN(d_2) + N(-d_1)$$

so that

$$s = -\ln[N(d_2) + N(-d_1)/L]/T$$

- 17.18** In this case $E_0 = 2$, $\sigma_E = 0.50$, $D = 5$, $r = 0.04$, and $T = 1$. Solving the simultaneous equations gives $V_0 = 6.80$ and $\sigma_V = 14.82$. The probability of default is $N(-d_2)$ or 1.15%.
- 17.19** At the end of each quarter for the first four years, the seller receives \$1 million. The seller pays \$70 million after four years and two months. The seller receives a final accrual payment of \$666,667.
- 17.20** A credit default swap insures a corporate bond issued by the reference entity against default. Its approximate effect is to convert the corporate bond into a risk-free bond. The buyer of a credit default swap has therefore chosen to exchange a corporate bond for a risk-free bond. This means that the buyer is long a risk-free bond and short a similar corporate bond.

- 17.21** Payoffs from credit default swaps depend on whether a particular company defaults. Arguably some market participants have more information about this than other market participants. (See Business Snapshot 17.2.)
- 17.22** Suppose that the principal is \$100. The present value of the bond if it were risk-free would be

$$2.5e^{-0.06 \times 0.5} + 2.5e^{-0.06 \times 1} + \dots + 2.5e^{-0.06 \times 4.5} + 102.5e^{-0.06 \times 5} = 95.3579$$

The present value of the expected loss from defaults is therefore $95.3579 - 90 = 5.3579$. The asset swap is structured so that the \$10 is paid initially. After that, \$2.50 is paid every six months. In return, floating plus a spread is received on the principal of \$100. The present value of the fixed payments is

$$10 + 2.5e^{-0.06 \times 0.5} + 2.5e^{-0.06 \times 1} + \dots + 2.5e^{-0.06 \times 4.5} + 102.5e^{-0.06 \times 5} = 105.3579$$

The spread over floating must therefore have a present value of 5.3579. The present value of \$1 received every six months for five years is 8.5105. The spread received every six months must therefore be $5.3579/8.5105 = \$0.6296$. The asset swap spread is therefore $2 \times 0.6296 = 1.2592\%$ per annum. This problem provides an illustration of the result in Problem 17.16.

Chapter 18

- 18.1** The new transaction will increase the bank's exposure to the counterparty if it tends to have a positive value whenever the existing transactions have a net positive value and a negative value whenever the existing transactions have a negative value. However, if the new transaction tends to offset the existing transactions, it is likely to have the incremental effect of reducing credit risk.
- 18.2** A company's own equity would not be good collateral. When the company defaults on the transactions it has with you, its equity is likely to be worth very little.
- 18.3** The statements in (a) and (b) are true. The statement in (c) is not. Suppose that ν_X and ν_Y are the exposures to X and Y. The expected value of $\nu_X + \nu_Y$ is the expected value of ν_X plus the expected value of ν_Y . The same is not true of 95% confidence limits.
- 18.4** Assume that defaults happen only at the end of the life of the forward contract. In a default-free world, the forward contract is the combination of a long European call and a short European put where the strike price of the options equals the delivery price and the maturity of the options equals the maturity of the forward contract. If the no-default value of the contract is positive at maturity, the call has a positive value and the put is worth zero. The impact of defaults on the forward contract is the same as that on the call. If the no-default value of the contract

is negative at maturity, the call has a zero value and the put has a positive value. In this case, defaults have no effect. Again, the impact of defaults on the forward contract is the same as that on the call. It follows that the contract has a value equal to a long position in a call that is subject to default risk and a short position in a default-free put.

- 18.5** The Black–Scholes–Merton price must be multiplied by $e^{-0.012 \times 3} = 0.964$. Black–Scholes–Merton overstates the price by about 3.6%.
- 18.6** If many contracts entered into by Company X are subject to the same default trigger, the effect of the default trigger may be to increase risk. If the company is downgraded so that the default trigger is activated, those of its counterparties who have contracts with negative values to Company X will request collateral. As a result, Company X is likely to experience liquidity problems and may be forced into bankruptcy.
- 18.7** A dealer is likely to have wrong-way risk when the counterparty is entering into a CDS selling credit protection to the dealer or when the counterparty is speculating. The dealer is likely to have right-way risk when the counterparty is entering into a CDS buying credit protection from the dealer or when the counterparty is partially hedging an existing position.
- 18.8** The cure period is the period that is assumed between an event of default and an early termination in CVA calculations.
- 18.9** All outstanding transactions between Party A and Party B are considered to be a single transaction if Party B defaults. It is therefore the total value of all outstanding transactions that determines the loss, if any, to Party A.
- 18.10** When a bank is experiencing financial difficulties, its probability of default increases. This increases the expected cost to its counterparties from a default by the bank. The latter is known as DVA and is an accounting entry that increases the value of transactions to the bank.
- 18.11** The part of CVA relating to the credit spread changes of counterparties is a component of market risk by regulators.
- 18.12** There will be losses in the CVA model if the value of the portfolio moves in the dealer's favor during a 10-day period immediately prior to the midpoint of a time interval. It is assumed that the dealer will fail to receive additional collateral from the counterparty.

Chapter 19

- 19.1** In Vasicek's model and Credit Risk Plus, a credit loss is recognized when a default occurs. In CreditMetrics, both downgrades and defaults lead to credit losses. In Vasicek's model, a Gaussian copula model of time to default is used. In Credit Risk Plus, a probability distribution is assumed for the default rate per year. In CreditMetrics, a Gaussian copula model is used to define rating transitions.

- 19.2** The constant level of risk assumption assumes that after a certain period of time, t , an instrument, X , if it has deteriorated is replaced with an instrument, Y , that has the same risk as X had originally. After a further period of time of length t , the instrument if it has deteriorated is again replaced with an instrument and has the same risk as X had originally, and so on.
- 19.3** The probability of an AAA-rated company being AAA at the end of two years is 80.79%. This is calculated by multiplying the transition matrix by itself. It incorporates the probability that it stays AAA (0.8985×0.8985), the probability that it will be downgraded to AA in the first year and then upgraded to AAA in the second year ($= 0.0935 \times 0.005$), etc. The probability that it will be AA at the end of the second year is 16.90%.
- 19.4** The probability of an AAA-rated company rating staying AAA over six months is 94.78%. The probability of it moving to Aa is 4.92%.
- 19.5** Movements in credit spreads for all companies over the next day could be assumed to be a random sample from their movements over the past 500 days. The disadvantages of this approach are that the companies have zero chance of defaulting and accurate daily credit spread data may not be available for all companies.
- 19.6** Using the binomial distribution, the probability of six or more defaults is 0.0005.
- 19.7** In this case, we must average the cumulative binomial distributions for 0.5% and 1.5% loss probabilities. The probability of six or more defaults is 0.0021. This shows that introducing some correlation increases the tail risk.
- 19.8** The autocorrelation is quite high at 0.492. This suggests that the credit VaR estimate should take account of recent default experience. If the default rate was high last year, it is more likely to be high this year.

Chapter 20

- 20.1** The definition includes all internal risks and external risks except those related to reputational risk and risks resulting from strategic decisions.
- 20.2** Based on the results reported by Shih, the loss would be $100 \times 3^{0.23}$ or \$128.7 million.
- 20.3** $\text{Prob}(v > x) = Kx^{-0.8}$. When $x = 20$ the probability is 0.1. This means that $K = 1.0986$. The probabilities of the specified losses being exceeded are (a) 5.74%, (b) 3.30%, and (c) 1.58%.
- 20.4** Moral hazard is handled by deductibles and by making premiums dependent on past claims. Adverse selection is handled by finding out as much as possible about a driver before insurance is granted and then modifying premiums as more information on the driver becomes available.
- 20.5** CEOs must prepare a statement asserting that the financial statements are accurate. They must return bonuses in the event that there is a restatement of financial statements.

- 20.6** If a trader operates within established risk limits and takes a loss, this is part of market risk. If risk limits are violated, the loss becomes classified as an operational risk.
- 20.7** (a) It is unlikely that an individual would not look after his or her health because of the existence of a life insurance contract. But it has been known for the beneficiary of a life insurance contract to commit murder to receive the payoff from the contract! (b) Individuals with short life expectancies are more likely to buy life insurance than individuals with long life expectancies.
- 20.8** External loss data are data relating to the losses of other banks. They are data obtained from sharing agreements with other banks or from data vendors. Vendor data are used to determine relative loss severity. They can be a useful indicator of the ratio of mean loss severity in Business Unit A to mean loss severity in Business Unit B or the ratio of the standard deviation of loss severity in Business Unit A to the standard deviation of loss severity in Business Unit B.
- 20.9** The Poisson distribution is often used for loss frequency. The lognormal distribution is often used for loss severity.
- 20.10** Examples of key risk indicators are staff turnover, number of failed transactions, number of positions filled by temps, ratio of supervisors to staff, number of open positions, and percentage of staff who did not take 10 days of consecutive leave in the past 12 months.
- 20.11** When the loss frequency is 3, the mean total loss is about 3.3 and the standard deviation is about 2.0. When the loss frequency is increased to 4, the mean loss is about 4.4 and the standard deviation of the loss is about 2.4.
- 20.12** The key inputs are the business indicator and (for banks with a business indicator over \$1 billion) the losses experienced over the previous 10 years. The business indicator is a measure of a bank's size that is relevant for operational risk purposes.
- 20.13** Each bank has an LC of $20 \times 15 = 300$. The ILM is

$$\ln \left(e - 1 + \frac{300}{200} \right)^{0.8} = 1.13$$

The capital of each bank is 200×1.13 or about 226 million euros.

Chapter 21

- 21.1** Investors did not know very much about the mortgages underlying the tranches that were created, and the waterfalls were complex.
- 21.2** The company or individual providing the quotes is prepared to buy at 50 and sell at 55. The mid-market quote is 52.5. The proportional bid-ask spread is $5.0/52.5$ or 0.0952.
- 21.3** The bid-ask spread for the holding in Company A is $0.01 \times 5,000 = \$50$. The bid-ask spread for the holding in Company B is $0.02 \times 3,000 = \$60$. The cost of unwinding the portfolio is $(50 + 60)/2$ or \$55.

- 21.4** The bid–ask spread for the first holding that we are 95% confident will not be exceeded is $5,000 \times (0.01 + 1.645 \times 0.01) = \132.24 . The bid–ask spread for the second holding that we are 95% confident will not be exceeded is $3,000 \times (0.02 + 1.645 \times 0.03) = \204.04 . The total cost of unwinding that we are 95% certain will not be exceeded is $(132.24 + 204.04)/2 = \$170.14$.
- 21.5** The amounts traded on successive days should be 15.9, 12.9, 10.0, 7.4, 5.2, 3.4, 2.2, 1.4, 0.9, and 0.7. The bid–ask spread cost is \$13.4. The total price variance is 36.6 so that the VaR for the market risk with a 95% confidence level is $1.645 \times \sqrt{36.6} = \9.9 . (The objective function that is minimized is the sum of these two, or \$23.3.)
- 21.6** LCR tests whether a bank has sufficient liquidity to survive 30 days of extreme stress (credit downgrades, loss of deposits, drawdowns on lines of credit, etc.). NSFR tests how well the maturities of assets and liabilities are matched and discourages overreliance on short-term sources of funds for long-term needs.
- 21.7** Wholesale deposits are more likely to disappear in stressed market conditions.
- 21.8** Their hedging led to a loss on the hedge and a gain on the position being hedged. The loss on the hedge gave rise to margin calls. Unfortunately, the position being hedged, although it increased in value to the company, was illiquid.
- 21.9** Positive feedback trading refers to situations where traders accentuate market movements. They buy when prices increase and sell when prices decrease. Negative feedback trading is when traders do the reverse; that is, they buy when prices decline and sell when prices increase. Positive feedback trading is liable to lead to a liquidity problem.
- 21.10** This is a VaR measure that includes an adjustment for the bid–ask spread costs that are incurred in a close–out of the position.
- 21.11** Liquidity black holes occur when most market participants want to be on one side of a market. Regulation is liable to lead to liquidity black holes because, when all financial institutions are regulated in the same way, they tend to want to respond to external economic events in the same way.
- 21.12** Liquidity black holes are typically caused by too many traders following the same trading strategy. If traders follow diverse trading strategies, liquidity black holes are less likely to occur.

Chapter 22

- 22.1** Marking to market involves valuing a position using the prices at which the same or similar positions are trading in the market. Marking to model occurs when a model plays a key role in determining the price of an instrument.
- 22.2** Models may have fundamental errors or they may be used inappropriately.
- 22.3** The model developer may leave. Also, other employees of the financial institution, as well as its regulators, can understand what has been done.

- 22.4** When plain vanilla call and put options are being priced, traders do use the Black–Scholes–Merton model as an interpolation tool. They calculate implied volatilities for the options that are actively traded. By interpolating between strike prices and between times to maturity, they estimate implied volatilities for other options. These implied volatilities are then substituted into the Black–Scholes–Merton model to calculate prices for these options. However, Black–Scholes–Merton is more than an interpolation tool when used for hedging.
- 22.5** 13.45%. We get the same answer by (a) interpolating between strike prices of 1.00 and 1.05 and then between maturities six months and one year and (b) interpolating between maturities of six months and one year and then between strike prices of 1.00 and 1.05.
- 22.6** The models of physics describe the behavior of physical processes. The models of finance ultimately describe the behavior of human beings.
- 22.7** It might notice that it is getting a large amount of business of a certain type because it is quoting prices different from its competitors. The pricing differences might also become apparent if it decides to unwind transactions and approaches competitors for quotes. Also, it might subscribe to a service where once a month it obtains the average price quotes by dealers for particular transactions.
- 22.8** Within-model hedging involves hedging against changes in variables that the model assumes to be stochastic. Outside-model hedging involves hedging against parameters that the model assumes to be constant.
- 22.9** P&L decomposition analyzes day-to-day profits and losses to distinguish the effects of (a) unhedged risks, (b) the hedging model operating imperfectly, and (c) new trades.
- 22.10** Level 1 transactions are transactions where there are quoted prices for the same transaction in active markets. Level 2 transactions are transactions where there are quoted prices for similar transactions in active markets or for the same transaction in markets that are not active. Level 3 transactions are transactions where some modeling assumptions are required.
- 22.11** Over-fitting involves trying to match too much of what is observed. Over-parameterization involves making a model too complex and difficult to use by introducing extra parameters.
- 22.12** An option with a strike price of 0.90 and a time to maturity of 1.5 years has an implied volatility of 14.85%. An option with a strike price of 0.95 and a time to maturity of 1.5 years has an implied volatility of 14.20%. An option with a strike price of 0.92 and a maturity of 1.5 years has an implied volatility of $0.6 \times 14.85\% + 0.4 \times 14.20\%$, which is 14.59%.
- 22.13** The term *marking to market* refers to the practice of revaluing instruments (usually daily) so that they are consistent with the market. The prices calculated for actively traded products do reflect market prices. The model is used merely as an interpolation tool. The term *marking to market* is therefore accurate for these products. The prices for structured products depend on the model being used, hence the term *marking to model*.

Chapter 23

- 23.1** It releases carbon dioxide into the atmosphere. This reduces the amount of heat that is radiated back into space.
- 23.2** The goal is to reduce temperature rises above pre-industrial levels in the 21st century to 1.5 degrees Celsius.
- 23.3** SSPs are Shared Socioeconomic Pathways. They are scenarios produced by IPCC estimating temperature increases under alternative assumptions.
- 23.4** A carbon tax is a tax on the tons of carbon dioxide emitted by a company. A cap and trade system caps the emissions from companies, but allows them to trade so that one company can sell some of its allowed emissions to another company.
- 23.5** Physical risks are the costs arising from the physical effect of global warming: heat waves, wild fires, hurricanes, etc. Transition risks are the costs associated with the move to lower emissions: closing plants, changing business practices, etc.
- 23.6** The physical risks associated with global warming must be carefully assessed. Life expectancy may decline. Insurance companies cannot average the cost of past risks to estimate the cost of future risks. They must extrapolate trends.
- 23.7** Environmental, Social, and Governance.
- 23.8** Greenwashing is the marketing of a company as environmentally friendly when it is not.

Chapter 24

- 24.1** In top-down approaches, risk limits are set at the senior management or board level for the whole organization. In bottom-up approaches, risk limits are set for the different business units. Both approaches are necessary in ERM because it is necessary to ensure that risk limits for the business units, when aggregated, are consistent with the risk limit for the whole organization.
- 24.2** The Bowman paradox is based on empirical data showing that taking higher risks tends to lead to lower returns as far as major strategic investments are concerned.
- 24.3** There is a danger that decisions will be made that are profitable in the short term but lead to long-term problems (e.g., negative publicity, loss of reputation, and lawsuits). Business Snapshots 5.4, 21.1, 21.2, and 21.3 are examples.
- 24.4** Part of the bonus for a year can be deferred with the possibility of clawbacks.
- 24.5** ERM takes a holistic (big picture) approach to risk management rather than separately considering the management of different types of risk (credit risk, market risk, operational risk, etc.).
- 24.6** In equation (24.1), we set $R = -0.2$, $R_F = 0.03$, $R_M = 0.1245$, $\sigma_M = 0.1688$, and $p = 0.9$. The appropriate beta is 1.89.

- 24.7** Each quarter from year 2 onward, the interest rate would increase by 4%. After 16 quarters, the rate would increase to $1.76\% + 16 \times 4\% = 65.76\%$.
- 24.8** Examples of cognitive biases are (a) anchoring (making too small adjustments to an initial estimate), (b) availability (recent information given too much weight), (c) representativeness (placing too much reliance on experience when it is limited or not representative of a current situation), and (d) inverting the conditionality.

Chapter 25

- 25.1** The removal of a competitor may be beneficial. However, banks enter into many contracts with each other. When one bank goes bankrupt, other banks are liable to lose money on the contracts they have with the bank. Also, other banks will be adversely affected if the bankruptcy reduces the public's overall level of confidence in the banking system.
- 25.2** Deposit insurance means that depositors are safe regardless of the risks taken by their financial institution. It is liable to lead to financial institutions taking more risks than they otherwise would because they can do so without the risk of losing deposits. This in turn leads to more bank failures and more claims under the deposit insurance system. Regulation requiring the capital held by a bank to be related to the risks taken is necessary to avoid this happening.
- 25.3** The credit risk on the swap is the risk that the counterparty defaults at some future time when the swap has a positive value to the bank.
- 25.4** The value of a currency swap is liable to deviate further from zero than the value of an interest rate swap because of the final exchange of principal. As a result, the potential loss from a counterparty default is higher.
- 25.5** There is some exposure. If the counterparty defaulted now, there would be no loss. However, interest rates could change so that at a future time the swap has a positive value to the financial institution. If the counterparty defaulted at that time, there would be a loss to the financial institution. The capital under Basel I would, from Table 25.2, be 0.5% of the swap's principal.
- 25.6** The risk-weighted assets for the three transactions are (a) \$1.875 million, (b) \$2 million, and (c) \$3 million, for a total of \$6.875 million. The capital required is 0.08×6.875 or \$0.55 million.
- 25.7** The net replacement ratio (NRR) is $2.5/4.5 = 0.556$. The credit equivalent amount is $2.5 + (0.4 + 0.6 \times 0.556) \times 9.25$ or \$9.28 million. The risk-weighted assets are \$4.64 million, and the capital required is \$0.371 million.
- 25.8** In this case, there is no value to the netting provisions.
- 25.9** This converts the estimated capital requirement to estimated risk-weighted assets. Capital required equals 8% of risk-weighted assets.
- 25.10** The trading book consists of instruments that are actively traded and marked to market daily. The banking book consists primarily of loans that are held to maturity

and not marked to market daily. The effect of the change is to move the client's borrowings from the banking book to the trading book. This typically reduces capital requirements. (However, the incremental risk charge in Basel II.5, which is discussed in Chapter 26, brings capital requirements back up to where they were before.)

- 25.11** Under Basel I, the capital charged for lending to a corporation is the same regardless of the credit rating of the corporation. This leads to a bank's return on capital being relatively low for lending to highly creditworthy corporations. Under Basel II, the capital requirements of a loan are tied much more carefully to the creditworthiness of the borrower. As a result, lending to highly creditworthy companies became more attractive.
- 25.12** Regulatory arbitrage involves entering into a transaction or series of transactions, without affecting the risks being taken, in order to reduce regulatory capital requirements.
- 25.13** EAD is the estimated exposure at default. LGD is the loss given default, which is the proportion of the exposure that will be lost if a default occurs. WCDR is the one-year probability of default in a bad year that occurs only one time in 1,000. PD is the probability of default in an average year. MA is the maturity adjustment, which allows for the fact that, in the case of instruments lasting longer than a year, there may be losses arising from a decline in the creditworthiness of the counterparty during the year as well as from a default during the year.
- 25.14** Under the simple approach, the risk weight of the counterparty is replaced by the risk weight of the collateral for the part of the exposure covered by the collateral. Under the comprehensive approach, the exposure is adjusted for possible increases and the collateral is adjusted for possible decreases in value. The counterparty's risk weight is applied to the excess of the adjusted exposure over the adjusted collateral.
- 25.15** The standardized approach uses external ratings to determine capital requirements (but in a more sophisticated way than in Basel I). In the IRB approach, the Basel II correlation model is used with PD being determined by the bank. In the advanced IRB approach, the Basel II correlation model is used with PD, LGD, EAD, and M being determined by the bank.
- 25.16** In the basic indicator approach, total capital is 15% of the average total annual gross income. In the standardized approach, gross income is calculated for different business lines, and capital as a percentage of gross income is different for different business lines. In the advanced measurement approach, the bank uses internal models to determine the 99.9% one-year VaR.
- 25.17** In this case, $\rho = 0.1216$, $WCDR = 0.0914$, and the capital requirement is $200 \times 0.7 \times (0.0914 - 0.01)$ or \$11.39 million. At least half of this must be Tier 1.
- 25.18** The probability of five or more exceptions is $1 - \text{BINOMDIST}(4, 250, 0.01, \text{TRUE})$ or 10.8%. It could be argued that regulators are using a confidence level of about 10% (rather than the more usual 5%) in choosing to reject a VaR model.

Chapter 26

- 26.1** The three major components of Basel II.5 are: the calculation of stressed VaR, a new incremental risk charge, and a comprehensive risk measure for instruments dependent on credit correlation.
- 26.2** The six major components of Basel III are: capital definitions and requirements, the capital conservation buffer, the countercyclical buffer, the leverage ratio, liquidity ratios, and counterparty credit risk.
- 26.3** VaR, as it is usually defined, is calculated from the most recent one to four years of daily data. Stressed VaR is calculated from a 250-day period in the past that would be particularly bad for the bank's current portfolio.
- 26.4** The incremental risk charge is calculated as the one-year 99.9% VaR for losses from credit instruments in the trading book. It takes account of rating changes and liquidity horizons. It was introduced because instruments in the trading book often attracted less capital than equivalent instruments in the banking book.
- 26.5** The capital requirements for the AAA-rated ABS are 1.6% of principal whereas the capital requirements for the AAA-rated ABS CDO are 3.2% of principal.
- 26.6** Tier 1 equity capital has increased from 2% to 7%, and the definition of equity capital has been tightened.
- 26.7** (a) 40%, (b) 20%.
- 26.8** In the leverage ratio, the denominator is not risk-weighted assets. It is total assets on the balance sheet without risk weighting plus derivatives exposures (calculated as in Basel I) and some off-balance-sheet items such as loan commitments.
- 26.9** The liquidity coverage ratio (LCR) is the ratio of high-quality liquid assets to net cash outflows during a stressed period of 30 days. The net stable funding ratio (NSFR) is the ratio of a weighted sum of the items on the "liabilities and net worth" side of the balance sheet divided by a weighted sum of the items on the "assets" side of the balance sheet.
- 26.10** The amount of stable funding would change to 81.6 and the NSFR would become $81.6/74.25 = 110\%$.
- 26.11** The credit value adjustment (CVA) is a charge to income reflecting expected losses from counterparty defaults in derivatives transactions. The new regulations require the exposure of CVA to credit spreads to be a component of market risk capital.
- 26.12** CoCo bonds are automatically converted into equity when a predefined trigger indicates that the bank's capital is low. They are attractive to banks because they do not affect return on equity prior to conversion. They are attractive to regulators because they are a source of capital that can absorb losses in stressed market conditions.

Chapter 27

- 27.1** In Basel I, market risk capital is calculated from the current 99% ten-day VaR, which is assumed to be $\sqrt{10}$ times the current 99% one-day VaR. In Basel II.5, there are two components of the capital. One is calculated from the current 99% VaR (as in Basel I); the other is calculated from the stressed 99% 10-day VaR. Each component is assumed to be $\sqrt{10}$ times the corresponding one-day VaR. In the FRTB, capital is calculated from the 97.5% stressed ES with the time horizon for a variable being dependent on its liquidity.
- 27.2** The 97.5% ES is $\mu + \sigma e^{-1.96^2/2}/(\sqrt{2\pi} \times 0.025) = \mu + 2.338\sigma$.
- 27.3** Consider how historical simulation works. The change from Day 0 to Day 10 and the change from Day 1 to Day 11 have nine days in common. They are therefore not independent. Indeed we would expect the two changes to be similar in most situations.
- 27.4** A theoretical advantage of ES over VaR is that it satisfies the subadditivity (diversification) condition in Section 11.5. A practical advantage is that it quantifies tail risk better.
- 27.5** The trading book consists of items that are held for trading and marked to market daily. The banking book consists of items that are expected to be held to maturity. Capital requirements are calculated in quite different ways for the two books, and regulators do not want to give banks a way of reducing capital by allowing them to choose the book they put an item in.
- 27.6** Credit spread risk is treated in the same way as other market risks. Jump-to-default risk is handled similarly to the credit risk associated with items in the banking book.

Chapter 28

- 28.1** Economic capital is a bank's own estimate of the capital it requires. Regulatory capital is the capital it is required to keep by bank supervisors.
- 28.2** The probability of a AA-rated company defaulting in one year.
- 28.3** Business risk includes risks relating to strategic decisions and reputation.
- 28.4** The models used for economic capital are likely to be broadly similar to those used to calculate regulatory capital in the case of market risk and operational risk. When calculating credit risk economic capital, a bank may consider it appropriate to use a different credit correlation model and different correlation parameters from those used in regulatory capital calculations.
- 28.5** The 99.97% worst-case value of the logarithm of the loss is $0.5 + 4 \times 3.43 = 14.23$. The 99.97% worst-case loss is therefore \$1.510 million. From the properties of the lognormal distribution, the expected loss is $\exp(0.5 + 4^2/2)$ or \$4,915. The capital requirement is therefore \$1.505 million.

- 28.6** The economic capital for Business Unit 1 is 96.85. The economic capital for Business Unit 2 is 63.87. The total economic capital is 124.66.
- 28.7** The incremental effect of Business Unit 1 on total economic capital is 60.78. The incremental effect of Business Unit 2 on total economic capital is 27.81. This suggests that $60.78/(60.78 + 27.81)$ or 68.61% of economic capital should be allocated to Business Unit 1, and $27.81/(60.78 + 27.81)$ or 31.39% to Business Unit 2. The marginal effect of increasing the size of Business Unit 1 by 0.5% is 0.4182. The marginal effect of increasing the size of Business Unit 2 by 0.5% is 0.2056. Euler's theorem is satisfied because the total economic capital is approximately equal to the sum of $0.4182/0.005$ and $0.2056/0.005$.
- 28.8** The capital is \$38 million and the return before tax is \$7 million. The before-tax RAROC is therefore 18.4%. In practice, the allocation of diversification benefits to this venture might reduce capital and increase RAROC.
- 28.9** RAROC can be used to compare the past performance of different business units (for performance appraisal) or to project the expected future performance of business units (to decide how capital should be allocated).

Chapter 29

- 29.1** Disintermediation occurs when two sides that previously interacted through an intermediary start to interact directly. Reintermediation occurs when, following disintermediation, a new technology intermediary (providing a different service from the previous intermediary) enters the market.
- 29.2** Machine learning is a branch of artificial intelligence where computers learn without being explicitly programmed. Distributed ledger technologies are technologies like blockchain where a database is consensually shared and synchronized across multiple computers.
- 29.3** Biometric authorization uses the unique biological characteristics of individuals to verify that they are who they say they are. Fingerprints, retinal scans, facial recognition, voice authentication, and heartbeat monitoring are examples of what can be used.
- 29.4** It might think a digital currency makes it easier to control the money supply and manage inflation.
- 29.5** In P2P lending, borrowers and lenders transact directly with each other using an online platform provided by a technology company. Equity crowdfunding is a mechanism that allows a broad group of investors to fund start-ups via an online platform.
- 29.6** It can be argued that regulation is a barrier to entry for FinTechs. In order to conform with regulation, FinTechs sometimes have to offer their services through an established financial institution.

- 29.7** It is true that if everyone invested only in index funds there would be no price discovery. We need some investors and fund managers to be following news reports and actively searching for stocks that should be bought or sold.
- 29.8** RegTech refers to technology developments that address regulatory compliance challenges. Examples are systems for online real-time reporting, systems to screen customers and transactions to avoid compliance issues, and online libraries making it easy for financial institutions to retrieve information about regulations on different subjects in different countries.
- 29.9** IBM recognized that it could not rely on the sale of mainframes and changed its culture and business model to reflect changes in the market. Kodak did not change its culture and business model to the same extent as the photography market changed. Banks will need to change both their culture and their business model in order to survive.

Glossary

ABS See Asset-Backed Security.

ABS CDO A security created from the tranches of different ABSs.

Accrued Interest The interest earned on a bond since the last coupon payment date.

Additional Tier 1 Capital Items such as non-cumulative preferred stock that do not qualify as Tier 1 equity capital.

Add-On Factor When the credit equivalent amount for a derivatives transaction is being calculated, this is the percentage of principal added to the current exposure to allow for possible future changes in the value of the derivative.

Adjoint Differentiation A way of modifying a computer program so that Greek letters are calculated in a computationally efficient way.

Advanced Measurement Approach The way in which the most sophisticated banks will be allowed to calculate regulatory capital for operational risk under Basel II.

Adverse Selection The phenomenon that, if an insurance company offers the same premiums to everyone, it tends to end up providing coverage for the worst risks.

Agency Costs Costs in a business relationship where the interests of the two parties are not perfectly aligned.

Alpha Return earned on a portfolio in excess of that predicted by the capital asset pricing model.

Alternative Investments See Hedge Funds.

American Option An option that can be exercised at any time during its life.

Analytic Result Result where the answer is in the form of an equation.

Arbitrage A trading strategy that takes advantage of two or more securities being mis-priced relative to each other.

Arbitrage Pricing Theory A theory where the return from an investment is assumed to depend on several factors.

Arbitrageur An individual engaging in arbitrage.

ASF Factor Available stable funding factor, a weighting factor used for sources of funding in the calculation of the Net Stable Funding Ratio.

Asian Option An option with a payoff dependent on the average price of the underlying asset during a specified period.

Ask Price The price that a dealer is offering to sell an asset.

Asked Price See Ask Price.

Asset Swap Exchanges the promised coupon on a bond for LIBOR plus a spread.

Asset-Backed Security A security created from the cash flows from bonds, mortgages, credit card receivables, or other instruments.

Asset Swap Exchanges the promised coupon on a bond for LIBOR plus a spread.

At-the-Money Option An option in which the strike price equals the price of the underlying asset.

Autocorrelation The correlation between the value of a variable and the value of the same variable k days later. (k is referred to as the time lag.)

Average Price Call Option An option giving a payoff equal to the greater of zero and the amount by which the average price of the asset exceeds the strike price.

Average Price Put Option An option giving a payoff equal to the greater of zero and the amount by which the strike price exceeds the average price of the asset.

Back-End Load Fee charged when an investment in a mutual fund is terminated.

Back Office Where the record keeping takes place.

Back-Testing Testing a value-at-risk or other model using historical data.

Backward Induction A procedure for working from the end of a tree to its beginning in order to value an option.

Banking Book Part of a bank's portfolio that consists of instruments that are expected to be held to maturity.

Bankruptcy Costs Costs such as lost sales, loss of key managers, and professional fees arising from a declaration of bankruptcy. These costs are not associated with the adverse events leading to bankruptcy.

Barrier Option An option whose payoff depends on whether the path of the underlying asset has reached a barrier (i.e., a certain predetermined level).

Basel I The first international agreement on the regulation of banks in 1988.

Basel II New international regulations for calculating bank capital introduced in 2007.

Basel II.5 Extra capital charges for items in the trading book, introduced following the credit crisis.

Basel III International banking regulations introduced in 2010 involving capital for the banking book and liquidity ratios.

Basel Committee The committee of bank regulators from around the world that produces standards intended to apply to banks in all countries.

Basic Indicator Approach The simplest way of calculating regulatory capital for operational risk under Basel II.

Basis The difference between the spot price and the futures price of a commodity.

- Basis Point** When used to describe an interest rate, a basis point is one-hundredth of 1% ($= 0.01\%$).
- Basis Risk** The risk to a hedger arising from uncertainty about the basis at a future time.
- Basket Credit Default Swap** A credit default swap where there are several reference entities.
- Basket Option** An option on a portfolio of assets.
- Bermudan Option** An option that can be exercised on certain dates during its life.
- Best Efforts** A phrase used to describe the situation where an investment bank agrees that it will do the best it can to sell a new issue of securities at a certain price, but does not guarantee that it will be able to sell them.
- Beta** A measure of the systematic risk of an asset.
- Bid-Ask Spread** The amount by which the ask price exceeds the bid price.
- Bid-Offer Spread** See Bid-Ask Spread.
- Bid Price** The price that a dealer is prepared to pay for an asset.
- Bilateral Clearing** An arrangement between two parties to handle transactions in the OTC market, often involving an ISDA master agreement.
- Binary Credit Default Swap** An instrument where there is a fixed dollar payoff in the event of a default by a particular company.
- Binary Option** An option with a discontinuous payoff; for example, a cash-or-nothing option or an asset-or-nothing option.
- Binomial Model** A model where the price of an asset is monitored over successive short periods of time. In each short period it is assumed that only two price movements are possible.
- Binomial Tree** A tree that represents how an asset price can evolve under the binomial model.
- BIS Accord** An agreement reached in 1988 between the central banks of 12 countries concerning how banks should be regulated.
- Bitcoin** The first cryptocurrency.
- Bivariate Normal Distribution** A joint distribution for two correlated variables, each of which has a normal distribution.
- Black's Model** An extension of the Black–Scholes–Merton model for valuing European options on futures contracts. It is used extensively in practice to value European options when the distribution of the asset price at maturity is assumed to be lognormal.
- Black–Scholes–Merton Model** A model for pricing European options on stocks, developed by Fischer Black, Myron Scholes, and Robert Merton.
- Blockchain** Distributed ledger technology where a secure database is consensually shared and synchronized across multiple computers.
- Bond Option** An option where a bond is the underlying asset.
- Bond Yield** The discount rate which, when applied to all the cash flows of a bond, causes the present value of the cash flows to equal the bond's market price.
- Bootstrap Method** A procedure for calculating the zero-coupon yield curve from market data. Also a statistical procedure for calculating confidence levels when distributions are determined empirically.

Bunching A tendency for days when the loss is greater than the value at risk to be clustered close together.

Business Risk When used by a bank, this refers to strategic risk (related to a bank's decision to enter new markets and develop new products) and reputation risk.

Buy and Hold A trading strategy where a position is set up and there is no subsequent trading of the position.

Calendar Days Includes every day.

Calibration A method for implying a model's parameters from the prices of actively traded options.

Callable Bond A bond containing provisions that allow the issuer to buy it back at a predetermined price at certain times during its life.

Call Option An option to buy an asset at a certain price by a certain date.

Cancelable Swap A swap that can be canceled by one side on prespecified dates.

Cap See Interest-Rate Cap.

Capital Adequacy The adequacy of the capital held by a bank or other financial institution.

Capital Asset Pricing Model A model relating the expected return on an asset to its beta.

Capital Conservation Buffer Extra equity capital that, under Basel III, must be kept so that it is available to absorb losses during downturns.

Caplet One component of an interest rate cap.

Cap Rate The rate determining payoffs in an interest rate cap.

Cascade Approach Describes the calculation methodology in FRTB's internal models approach.

Case-Shiller Index An index of house prices in the United States.

Cash-Flow Mapping A procedure for representing an instrument as a portfolio of zero-coupon bonds for the purpose of calculating value at risk.

Cash Settlement A procedure for settling a contract in cash rather than by delivering the underlying asset.

CAT Bond A bond where the interest and, possibly, the principal paid are reduced if a particular category of "catastrophic" insurance claims exceeds a certain amount.

CCAR The Comprehensive Capital Analysis and Review is a stress test in the United States for banks with consolidated assets over \$50 billion.

CCP See Central Clearing Party.

CDD See Cooling Degree Days.

CDO See Collateralized Debt Obligation.

CDO Squared An instrument in which the default risks in a portfolio of CDO tranches are allocated to new securities.

CDS See Credit Default Swap.

CDS-Bond Basis The excess of the CDS spread over the asset swap spread.

CDX An index of the credit quality of 125 North American investment grade companies.

Central Clearing The use of a central counterparty (CCP) for OTC transactions.

- Central Counterparty** Clearing house for over-the-counter transactions.
- Central Limit Theorem** A theorem that states that the sum of a large number of independent identically distributed variables is approximately normally distributed even if the variables themselves are not normally distributed.
- Cheapest-to-Deliver Bond** A bond that is cheapest to buy and then deliver in a futures contract or a credit default swap contract.
- Chinese Walls** A phrase used to describe the policies in place within a financial institution that prevent information flowing from one part of the financial institution to another when this would be disadvantageous to one or more of the financial institution's clients.
- Cholesky Decomposition** A method of sampling from a multivariate normal distribution.
- Clawback** In the case of a hedge fund, a clawback allows a percentage of previous incentive fees to be recouped to compensate investors for current losses. In the case of a deferred bonus, it leads to the bonus not being paid if financial performance deteriorates.
- Clean Price of Bond** The quoted price of a bond. The cash price paid for the bond (or dirty price) is calculated by adding the accrued interest to the clean price.
- Clearing House** A firm that guarantees the performance of the parties in an exchange-traded derivatives transaction. (Also referred to as a clearing corporation.)
- Clearing Margin** A margin posted by a member of a clearing house.
- Closed-End Fund** A mutual fund where there is a fixed number of shares.
- CoCo** See Contingent Convertible Bond.
- Cognitive Bias** Illogical pattern of judgment.
- Coherent Risk Measure** A risk measure satisfying a number of conditions.
- Collar** See Interest-Rate Collar.
- Collateral** Cash or marketable securities that must be posted by a party to a transaction.
- Collateralization** A system for posting collateral by one or both parties in a derivatives transaction.
- Collateralized Debt Obligation** A way of packaging credit risk. Several classes of securities (known as tranches) are created from a portfolio of bonds and there are rules for determining how the costs of defaults are allocated to classes.
- Commercial Bank** A bank that takes deposits and makes loans.
- Component Value at Risk** VaR corresponding to a component of a portfolio. Defined so that the sum of the component VaRs for the components of a portfolio equals the VaR for the whole portfolio.
- Compounding Frequency** This defines how an interest rate is measured.
- Compounding Swap** A swap where interest compounds instead of being paid.
- Compound Option** An option on an option.
- Comprehensive Risk Measure** A risk measure that calculates the capital charge for instruments dependent on credit correlation.
- Conditional Tail Expectation** See Expected Shortfall.

Conditional Value at Risk (C-VaR) See Expected Shortfall.

Confidence Level A probability indicating the level of confidence that an event will (or will not) happen. In value at risk calculations, the confidence level is the probability that the loss will not exceed the value at risk.

Confirmation A contract confirming a verbal agreement between two parties to a trade in the over-the-counter market.

Constant Level of Risk An assumption in the calculation of a risk measure that a financial institution will trade periodically so that its risk at the beginning of each period is the same as today. (Can be contrasted with Buy and Hold.)

Consumption Asset An asset held for consumption rather than investment.

Contagion See Credit Contagion.

Contingent Convertible Bond (CoCo) Bond that is automatically converted to equity when there is a trigger indicating that the issuer needs more equity capital.

Continuous Compounding A way of quoting interest rates. It is the limit as the assumed compounding interval is made smaller and smaller.

Convenience Yield A measure of the benefits from ownership of an asset that are not obtained by the holder of a long futures contract on the asset.

Conversion Factor A factor multiplied by principal to convert an off-balance-sheet item to its credit equivalent amount.

Convertible Bond A corporate bond that can be converted by the holder into a predetermined amount of the company's equity at certain times during its life.

Convexity A measure of the curvature in the relationship between bond prices and bond yields.

Convexity Adjustment An overworked term. For example, it can refer to the adjustment necessary to convert a futures interest rate to a forward interest rate. It can also refer to the adjustment to a forward rate that is sometimes necessary when instruments are valued.

Cooke Ratio A ratio of capital to risk-weighted assets under Basel I.

Cooling Degree Days The maximum of zero and the amount by which the daily average temperature is greater than 65 degrees Fahrenheit. The average temperature is the average of the highest and lowest temperatures (midnight to midnight).

Copula A way of defining the correlation between variables with known distributions.

Core Tier 1 Capital See Tier 1 Equity Capital.

Cornish-Fisher Expansion An approximate relationship between the fractiles of a probability distribution and its moments.

Correlation A measure of the extent to which there is a linear relation between two variables.

Correlation Matrix An $n \times n$ matrix where the $\{i, j\}$ element is the correlation between variable i and variable j .

Cost of Carry The storage costs plus the cost of financing an asset minus the income earned on the asset.

Countercyclical Buffer An extra capital charge, under Basel III, that is left to the discretion of national regulators.

- Counterparty** The opposite side in a financial transaction.
- Coupon** The interest payment made on a bond.
- Covariance** A measure of the linear relationship between two variables (equals the correlation between the variables times the product of their standard deviations).
- Covariance Rate** The covariance between daily returns of two variables.
- Covered Call** A short position in a call option on an asset combined with a long position in the asset.
- Crashophobia** The fear of a stock market crash similar to that in 1987 that some people claim causes market participants to increase the value of deep-out-of-the-money put options.
- Credit Contagion** The process whereby credit losses in one sector of the world economy lead to credit losses in other sectors.
- Credit Default Swap** An instrument that gives the holder the right to sell a bond for its face value in the event of a default by the issuer.
- Credit Derivative** A derivative whose payoff depends on the creditworthiness of one or more companies or countries.
- Credit Equivalent Amount** Size of loan that is considered equivalent to an off-balance-sheet transaction in Basel I.
- Credit Event** A default or other event that triggers a payout in a credit default swap.
- Credit Indices** Indices that track the cost of buying protection for companies in a specified portfolio.
- CreditMetrics** A procedure for calculating credit value at risk.
- Credit Migration** The movement of a company from one rating category to another.
- Credit Rating** A measure of the creditworthiness of a bond issue.
- Credit Ratings Transition Matrix** A table showing the probability that a company will move from one credit rating to another during a certain period of time.
- Credit Risk** The risk that a loss will be experienced because of a default by the counterparty in a derivatives transaction.
- Credit Risk Plus** A procedure for calculating credit value at risk.
- Credit Spread** The cost of buying credit protection. Also, the difference between the yield on a bond that might default and the yield on a risk-free bond.
- Credit Support Annex** Part of an ISDA master agreement specifying collateral arrangements.
- Credit Transition Matrix** See Credit Ratings Transition Matrix.
- Credit Value Adjustment** The adjustment to the value of derivatives transactions made by a dealer to allow for the possibility that the counterparty may default.
- Credit Value at Risk** The credit loss that will not be exceeded at some specified confidence level.
- Crowdfunding** Funding a project or a company using an online platform.
- Cryptocurrency** Digital currency designed to provide a secure medium of exchange.
- CSA** See Credit Support Annex.
- Cumulative Distribution Function** The probability that a variable will be less than x as a function of x .

Cure Period A period of time assumed to elapse between a default event and a termination in OTC derivatives transactions that are settled bilaterally.

Currency Swap A swap where interest and principal in one currency are exchanged for interest and principal in another currency.

Current Exposure Method A method for calculating exposure at default on derivatives in Basel I and Basel II. It sets exposure at default equal to current exposure plus an add-on factor.

Current VaR or ES VaR or ES calculated using most recent data, not data from a stressed period.

Curvature Gamma.

CVA See Credit Value Adjustment.

Day Count A convention for quoting interest rates.

Day Trade A trade that is entered into and closed out on the same day.

Debit Value Adjustment Adjustment to the value of derivatives transactions made by a dealer to allow for the possibility that the dealer may default.

Debt Value Adjustment See Debit Value Adjustment.

Default Correlation Measures the tendency of two companies to default at about the same time.

Default Intensity See Hazard Rate.

Deferred Annuity An annuity where time elapses between the payment of the funds that will provide the annuity and the start of the annuity.

Defined Benefit Plan A pension plan where there is a formula defining the pension that will be received. Typically, the formula depends on the number of years of service and the salary during the final years of service.

Defined Contribution Plan A pension plan where an employee's contributions plus the contributions made for the employee by the employer are kept in a separate account and invested. On retirement the funds in the account are usually converted into an annuity. Sometimes, they can be taken out as a lump sum.

Deleveraging The process of individuals and companies reducing their borrowing.

Delivery Price The price that will be paid or received in a forward contract.

Delta The rate of change of the price of a derivative with the price of the underlying asset.

Delta Hedging A hedging scheme that is designed to make the price of a portfolio of derivatives insensitive to small changes in the price of the underlying asset.

Delta-Neutral Portfolio A portfolio with a delta of zero so that there is no sensitivity to small changes in the price of the underlying asset.

Dependence Variable A is dependent on variable B if knowing the value of B affects the probability density function of A.

Deposit Insurance Government programs for providing restitution to the depositors of a bank in the event that the bank fails.

DerivaGem Software for valuing options, available on the author's website.

Derivative An instrument whose price depends on, or is derived from, the price of another asset.

- Deterministic Variable** A variable whose future value is known.
- DFAST** The Dodd–Frank Act Stress Test is a stress test in the United States for banks with consolidated assets over \$10 billion.
- Digital Currency** Type of currency available only in digital form. It is not available as banknotes or coins.
- Directed Brokerage** A phrase used to describe the situation where a fund directs its trades to a brokerage house in return for the brokerage house advising its clients to buy the fund.
- Dirty Price of Bond** The cash price of a bond.
- Discount Bond** See Zero-Coupon Bond.
- Discount Instrument** An instrument, such as a Treasury bill, that provides no coupons.
- Discount Rate** The annualized dollar return on a Treasury bill or similar instrument expressed as a percentage of the final face value.
- Disintermediation** Change where less use is made of intermediaries.
- Distance to Default** The number of standard deviations that the value of a company's assets must move for a default to be triggered.
- Diversification** Reducing risk by dividing a portfolio between many different assets.
- Dividend** A cash payment made to the owner of a stock.
- Dividend Yield** The dividend as a percentage of the stock price.
- Dodd–Frank Act** An act in the United States that was introduced following the credit crisis and is designed to protect consumers and investors, avoid future bailouts, and monitor the functioning of the financial system more carefully.
- Dollar Convexity** The convexity of an interest rate–dependent portfolio multiplied by the value of the portfolio.
- Dollar Duration** The duration of an interest rate–dependent portfolio multiplied by the value of the portfolio.
- Down-and-In Option** An option that comes into existence when the price of the underlying asset declines to a prespecified level.
- Down-and-Out Option** An option that ceases to exist when the price of the underlying asset declines to a prespecified level.
- Downgrade Trigger** A clause in a contract that states that the contract can be terminated, or collateral demanded, by one side if the credit rating of the other side falls below a certain level.
- Duration** A measure of the average life of a bond. It is also an approximation of the ratio of the proportional change in the bond price to the absolute change in its yield.
- Duration Matching** A procedure for matching the durations of assets and liabilities.
- Dutch Auction** A process where investors provide bids indicating the number of shares they are prepared to buy and the price they are prepared to pay. Suppose there are N shares on offer. The price that clears the market is the highest price, P , that is such that investors are prepared to buy more than N shares at either P or a higher price than P . The investors who have bid more than P get their orders filled at a price of P . The investors who have bid P get part of their orders filled at P .

DVA See Debit Value Adjustment.

DV01 The impact of a one-basis-point increase in all interest rates.

Dynamic Hedging A procedure for hedging an option position by periodically changing the position held in the underlying asset. The objective is usually to maintain a delta-neutral position.

Dynamic Scenarios Scenarios involving adverse movements in market variables that consider the way companies will respond to the adverse movements.

EAD See Exposure at Default.

Early Exercise Exercise prior to the maturity date.

Early Termination Event The early termination of OTC derivative transactions by one side because of an event of default by the other side.

Economic Capital The capital that a bank's own calculation indicates it needs.

Efficient Frontier The optimal trade-offs for an investor between expected return and standard deviation of return.

Efficient Market Hypothesis A hypothesis that asset prices reflect relevant information.

Electronic Trading System of trading where a computer is used to match buyers and sellers.

Embedded Option An option that is an inseparable part of another instrument.

Empirical Research Research based on historical market data.

Endowment Life Insurance Insurance that pays out a predetermined amount on the maturity of the contract or on the policyholder's death, whichever is earlier.

Enterprise Risk Management A holistic approach to risk management.

Equity Swap A swap where the return on an equity portfolio is exchanged for either a fixed or a floating rate of interest.

Equity Tranche The tranche that is first to be affected by losses on the underlying portfolio.

ERM See Enterprise Risk Management.

ES See Expected Shortfall.

Eurocurrency A currency that is outside the formal control of the issuing country's monetary authorities.

Eurodollar A dollar held in a bank outside the United States.

Eurodollar Futures Contract A futures contract written on a Eurodollar deposit.

Eurodollar Interest Rate The interest rate on a Eurodollar deposit.

European Option An option that can be exercised only at the end of its life.

Event of Default An event such as a bankruptcy or a failure to make payments as they become due or a failure to post collateral as it becomes due.

EWMA Exponentially weighted moving average.

Exception A situation in back-testing where realized loss exceeds the value-at-risk estimate.

Excess Cost Layers The costs for an insurance company on a particular type of business that are within a certain range.

Excess Kurtosis A measure of the extent to which the tails of a probability distribution are fatter than those of a normal distribution.

- Excess Spread** A situation where the aggregate return promised to the tranches is less than the promised return on the underlying assets.
- Exchange-Traded Fund** A fund created in such a way that units can be exchanged for the underlying shares and vice versa by institutional investors.
- Exchange-Traded Market** A market organized by an exchange such as the New York Stock Exchange or Chicago Board Options Exchange.
- Ex-Dividend Date** When a dividend is declared, an ex-dividend date is specified. Investors who own shares of the stock just before the ex-dividend date receive the dividend.
- Exercise Price** The price at which the underlying asset may be bought or sold in an option contract. (Also called the strike price.)
- Exotic Option** A nonstandard option.
- Expectations Theory** The theory that forward interest rates equal expected future spot interest rates.
- Expected Shortfall** Expected loss during N days conditional on being in the $(100 - X)\%$ tail of the distribution of losses. The variable N is the time horizon and $X\%$ is the confidence level.
- Expected Tail Loss** See Expected Shortfall.
- Expected Value of a Variable** The average value of the variable obtained by weighting the alternative values by their probabilities.
- Expense Ratio** Ratio of expenses to value of assets under management for a fund.
- Expiration Date** The end of life of a contract.
- Exponentially Weighted Moving Average Model** A model to provide forecasts for a variable from historical data where the weights applied to data decline exponentially as we move back through time. It is sometimes applied to variances and covariances in value at risk calculations.
- Exponential Weighting** A weighting scheme where the weight given to an observation depends on how recent it is. The weight given to an observation t time periods ago is λ times the weight given to an observation $t - 1$ time periods ago where $\lambda < 1$.
- Exposure at Default** The maximum amount that could be lost (assuming no recovery) when a default occurs.
- Extreme Value Theory** A theory enabling the shape of the tails of a distribution to be estimated from data.
- Factor** Source of uncertainty.
- Factor Analysis** An analysis aimed at finding a small number of factors that describe most of the variation in a large number of correlated variables. (Similar to a principal components analysis.)
- Factor Copula** A copula involving several variables where a factor model describes the correlation structure of the transformed variables.
- Factor Loadings** The values of variables in a factor model when we have one unit of a particular factor and no units of other factors.
- Factor Model** A model where a set of correlated variables are assumed to depend linearly on a number of uncorrelated factors.

Factor Score In a factor model, this is the amount of a factor present in a particular observation.

Fair Value Accounting Involves showing financial instruments on the balance sheet at their market value.

FICO A credit score developed by Fair Isaac Corporation.

Financial Intermediary A bank or other financial institution that facilitates the flow of funds between different entities in the economy.

FinTech Application of technology in finance.

Firm Commitment Phrase used to describe the situation where an investment bank guarantees that a new issue of securities will be sold at a certain price. If investors do not want to buy the securities, the investment bank will have to do so.

Fixed Annuity An annuity (usually for life) where payments that will be received by the policyholder are fixed in advance.

Floor See Interest-Rate Floor.

Floor-Ceiling Agreement See Collar.

Floorlet One component of a floor.

Floor Rate The rate in an interest rate floor agreement.

Foreign Currency Option An option on a foreign exchange rate.

Forward Contract A contract that obligates the holder to buy or sell an asset for a predetermined delivery price at a predetermined future time.

Forward Exchange Rate The forward price of one unit of a foreign currency.

Forward Interest Rate The interest rate for a future period of time implied by the rates prevailing in the market today.

Forward Price The delivery price in a forward contract that causes the contract to be worth zero.

Forward Rate Can refer to a forward interest rate or a forward exchange rate.

Forward Rate Agreement (FRA) An agreement that a certain interest rate will apply to a certain principal amount for a certain time period in the future.

Front-End Load A fee charged when a mutual fund is purchased.

Front Office Where the trading takes place.

Front Running Trading a stock on a personal account before a fund trades a large volume of the stock.

FRTB See Fundamental Review of the Trading Book.

Fundamental Review of the Trading Book A new way of assigning capital for market risk proposed by the Basel Committee.

Fund of Funds A fund that invests in a portfolio of different hedge funds.

Futures Contract A contract that obligates the holder to buy or sell an asset at a predetermined delivery price during a specified future time period. The contract is settled daily.

Futures Option An option on a futures contract.

Futures Price The delivery price currently applicable to a futures contract.

G-30 Policy Recommendations A set of recommendations concerning derivatives issued by nonregulators in 1993.

- Gamma** The rate of change of delta with respect to the asset price.
- Gamma-Neutral Portfolio** A portfolio with a gamma of zero.
- Gap Management** A procedure for matching the maturities of assets and liabilities.
- GARCH Model** A model for forecasting volatility where the variance rate follows a mean-reverting process.
- Gaussian Copula Model** A copula model based on the multivariate normal distribution.
- Girsanov's Theorem** A theorem showing that when we move from a world with one set of risk preferences to a world with another set of risk preferences the drifts of stochastic variables may change but their volatilities remain the same.
- Glass-Steagall Act** An act passed in the United States separating commercial and investment banks.
- Greek Letters** See Greeks.
- Greeks** Hedge parameters such as delta, gamma, vega, theta, and rho.
- Gross Income** Net interest income plus non-interest income.
- Group Life Insurance** Life insurance arranged for a group of people, usually by an employer.
- G-SIB** Global systemically important bank.
- Haircut** The discount applied to the value of an asset when it is used as collateral.
- Hazard Rate** Measures probability of default in a short period of time conditional on no earlier default.
- HDD** See Heating Degree Days.
- Heating Degree Days** The maximum of zero and the amount by which the daily average temperature is less than 65 degrees Fahrenheit. The average temperature is the average of the highest and lowest temperatures (midnight to midnight).
- Hedge** A trade designed to reduce risk.
- Hedge Funds** Funds that are subject to less restrictions and less regulation than mutual funds, but are limited in the extent to which they can sell units in the fund to the general public.
- Hedger** An individual who enters into hedging trades.
- Hedge Ratio** The ratio of the size of a position in a hedging instrument to the size of the position being hedged.
- High-Frequency Trading (HFT)** Trading generated by the use of computer algorithms.
- High-Water-Mark Clause** Clause stating that gains equal to previous losses must be made before incentive fees apply.
- Historical Default Probability** Default probability estimated from historical default experience.
- Historical Simulation** A simulation based on historical data.
- Historical Volatility** A volatility estimated from historical data.
- Holiday Calendar** Calendar defining which days are holidays for the purposes of determining payment dates in a financial transaction.
- Hurdle Rate** The return that must be earned by a hedge fund before the incentive fee applies.

- Hybrid Approach** An approach to aggregating different types of economic capital.
- IFRS 9** A new accounting standard requiring banks to value loans by taking expected default losses into account.
- Implied Default Probability** See Risk-Neutral Default Probability.
- Implied Volatility** Volatility implied from an option price using the Black–Scholes–Merton or a similar model.
- Inception Profit** Profit created by selling a derivative for more than its theoretical value.
- Incremental Risk Charge** An extra capital charge for credit-sensitive instruments in the trading book. The charge reflects the fact that capital requirements for credit instruments in the banking book are often higher than for equivalent-risk instruments in the trading book.
- Incremental Value at Risk** The difference between the value at risk with and without a particular component of the portfolio.
- Independence** Variable A is independent of variable B if knowing the value of variable B does not affect the probability density function for variable A.
- Independent Amount** Collateral required by a dealer in bilateral OTC trading that is independent of the value of outstanding transactions.
- Index Fund** A fund that is designed to match the performance of a particular stock index.
- Initial Margin** The cash required from a futures trader at the time of the trade.
- Initial Public Offering** An offering to investors of shares in a company for the first time. Prior to an IPO, shares are typically held only by the company's founders, its employees, and the providers of venture capital.
- Instantaneous Forward Rate** A forward rate for a very short period of time in the future.
- Interest-Rate Cap** An option that provides a payoff when a specified interest rate is above a certain level. The interest rate is a floating rate that is reset periodically.
- Interest-Rate Collar** A combination of an interest-rate cap and an interest-rate floor.
- Interest-Rate Derivative** A derivative whose payoffs are dependent on future interest rates.
- Interest-Rate Floor** An option that provides a payoff when an interest rate is below a certain level. The interest rate is a floating rate that is reset periodically.
- Interest-Rate Option** An option where the payoff is dependent on the level of interest rates.
- Interest-Rate Swap** An exchange of a fixed rate of interest on a certain notional principal for a floating rate of interest on the same notional principal.
- Internal Credit Rating** A credit rating produced by a financial institution rather than by an outside agency such as Moody's or Standard & Poor's.
- In-the-Money Option** Either (a) a call option where the asset price is greater than the strike price or (b) a put option where the asset price is less than the strike price.
- Intrinsic Value** For a call option, this is the greater of the excess of the asset price over the strike price and zero. For a put option, it is the greater of the excess of the strike price over the asset price and zero.

- Investment Asset** An asset held by a significant number of individuals for investment purposes.
- Investment Bank** A bank that helps companies issue debt and equity securities.
- Investment Grade** A bond or other instrument with a credit rating of BBB (Baa) or above.
- IPO** See Initial Public Offering.
- IRB Approach** Internal Ratings Based approach for assessing credit risk capital in Basel II.
- IRC** See Incremental Risk Charge.
- ISDA** International Swaps and Derivatives Association.
- ISDA Master Agreement** Agreement between two parties governing their OTC derivatives trades.
- Isserlis' Theorem** A theorem for calculating the mean of the product of zero-mean normally distributed variables.
- ITrxxx** An index of the credit quality of 125 European investment-grade companies.
- Jump-to-Default Risk** The risk that a company will default. It can be contrasted with credit spread risk, which is the risk that the company's credit spread will increase without the company defaulting.
- Junk Bond** A non-investment-grade bond.
- Key Risk Indicators** Indicators to track the level of operational risk.
- Kurtosis** A measure of the heaviness of the tails of a distribution.
- Late Trading** Trading a mutual fund after 4 P.M. at the 4 P.M. price.
- Leveraging** Individuals and companies increasing their borrowings.
- LGD** See Loss Given Default.
- Liar Loan** A loan where the borrower did not tell the truth on the application form.
- LIBOR** See London Interbank Offered Rate.
- LIBOR-in-Arrears Swap** A swap where the interest paid on a date is determined by the interest rate observed on that date (not by the interest rate observed on the previous payment date).
- LIBOR-OIS Spread** The difference between the LIBOR rate of interest and the OIS rate of interest for a particular maturity.
- LIBOR/Swap Zero Curve** Zero rates that are calculated from LIBOR rates, eurodollar futures, and swap rates.
- LIBOR Zero Curve** See LIBOR/Swap Zero Curve.
- Life Insurance** Insurance where payouts depend on when a person dies.
- Linear Model** A model where the change in the portfolio value depends linearly on the returns of the underlying market variables.
- Linear Product** A derivative product whose price depends linearly on one or more underlying variables.
- Liquidity-Adjusted VaR** A value at risk calculation that takes account of the impact of the bid-offer spread when positions are unwound.
- Liquidity Black Hole** A situation where liquidity dries up because everyone wants to be on the same side of the market.

Liquidity Coverage Ratio The ratio of high-quality liquid assets to cash outflows during a stressed period of 30 days.

Liquidity Funding Risk The risk that sources of funding will dry up.

Liquidity Horizon The time horizon used for a risk factor in FRTB's expected short-fall calculations.

Liquidity Preference Theory A theory leading to the conclusion that forward interest rates are above expected future spot interest rates.

Liquidity Premium The amount that forward interest rates exceed expected future spot interest rates.

Liquidity Trading Risk The risk that it will not be possible to unwind positions in financial instruments at their theoretical price.

Living Will A plan for winding up a financial institution in an orderly way so that some parts possibly survive.

Lognormal Distribution A variable has a lognormal distribution when the logarithm of the variable has a normal distribution.

London Interbank Offered Rate The rate at which an AA-rated bank can borrow from other banks.

Longevity Bond A bond where a population of individuals is defined and the coupon paid at a particular time is proportional to the number of people in the population who are alive at that time.

Longevity Risk The risk that people will live longer than expected.

Long Position A position involving the purchase of an asset.

Long-Tail Risk The risk in property-casualty insurance that there will be claims made well after the end of the life of a policy.

Lookback Option An option whose payoff is dependent on the maximum or minimum of the asset price achieved during a certain period.

Loss Given Default The percentage of the exposure to a counterparty that is lost when a default by the counterparty occurs.

Macaulay's Duration The weighted average of the times when cash flows are received where the weight applied to a cash flow is proportional to the present value of the cash flow.

Machine Learning A branch of artificial intelligence that allows computers to learn without being explicitly programmed.

Maintenance Margin When the balance in a trader's margin account falls below the maintenance margin level, the trader receives a margin call requiring the margin account to be topped up.

Margin The amount of collateral required in cash or marketable securities from a trader.

Marginal Distribution The unconditional distribution of a variable.

Marginal Value at Risk The rate of change of the value at risk with the size of one component of the portfolio.

Margin Call A request for extra margin when the balance in the margin account falls below the maintenance margin level.

Margin Period of Risk See Cure Period.

- Marginal Distribution** The unconditional distribution of a variable.
- Market Maker** A trader who is willing to quote both bid and offer prices for an asset.
- Market Model** A model most commonly used by traders.
- Market Portfolio** A portfolio consisting of the universe of all possible investments.
- Market Risk** A risk relating to movements in market variables.
- Market Timing** Frequent trading of a mutual fund to benefit from stale prices or to do late trading.
- Marking to Market** The practice of revaluing a financial instrument to reflect the current values of the relevant market variables.
- Marking to Model** Use of a model to determine the current value of a financial instrument.
- Maturity Date** The end of the life of a contract.
- Maximum Likelihood Method** A method for choosing the values of parameters by maximizing the probability of a set of observations occurring.
- Mean Reversion** The tendency of a market variable (such as a volatility or an interest rate) to revert back to some long-run average level.
- Merton's Model** A model that uses equity prices to estimate default probabilities. (Other models developed by Merton are also sometimes referred to as Merton's model.)
- Mezzanine Tranche** A tranche that experiences losses after the equity tranche but before the senior tranche.
- Middle Office** Where risk management takes place.
- Minimum Transfer Amount** The minimum amount transferred in a collateralization agreement.
- Model Building Approach** The use of a model to estimate value at risk.
- Model Risk** The risk relating to the use of models to price derivative products.
- Modified Duration** A modification to the standard duration measure so that it more accurately describes the relationship between proportional changes in a bond price and actual changes in its yield. The modification takes account of the compounding frequency with which the yield is quoted.
- Money Center Banks** Banks with a global presence that fund themselves to a significant extent in wholesale markets.
- Monte Carlo Simulation** A procedure for randomly sampling changes in market variables.
- Moral Hazard** The possibility that the behavior of an insured entity will change because of the existence of an insurance contract.
- Mortality Risk** The risk that people will die earlier than expected.
- Multibank Holding Company** A holding company that owns several banks. This structure was used in the United States for getting around some bank regulations.
- Multivariate Normal Distribution** The joint distribution of many variables, each of which is normal.
- Mutual Fund** A vehicle for investing the funds of many small investors.

Naked Position A short position in a call option that is not combined with a long position in the underlying asset.

National Association of Insurance Commissioners An national body that provides services to the insurance regulators in each state in the United States.

Negative Feedback Trading Trading where assets are sold after a price increase and bought after a price decrease.

Net Asset Value The value of a fund's investments divided by the number of shares in the fund.

Net Interest Income The excess of interest earned over interest paid for a bank.

Net Replacement Ratio The ratio of current exposure with netting to current exposure without netting.

Net Stable Funding Ratio The weighted average of funding available divided by weighted average of funding required.

Netting The ability to offset contracts with positive and negative values in the event of a default by a counterparty.

Neural Networks A machine learning tool that allows computers to recognize patterns in a similar way to humans.

NINJA A term used to describe a poor credit risk: no income, no job, no assets.

Node Delta When the term structure is piecewise linear, this measures the impact on a portfolio of changing the interest rate corresponding to one of the vertices by one basis point while keeping the other vertices unchanged.

Non-Investment Grade A bond or other instrument with a credit rating below BBB (Baa).

Nonlife Insurance See Property–Casualty Insurance.

Nonlinear product A derivative product that is not linearly dependent on the underlying variables.

Nonperforming Loan A loan where interest is more than 90 days overdue.

Nonsystematic Risk Risk that can be diversified away.

Normal Distribution The standard bell-shaped distribution of statistics.

Normal Market A market where futures prices increase with maturity.

Notional Principal The principal used to calculate payments in a derivatives transaction. The principal is notional because it is neither paid nor received.

Numerical Procedure A method of calculation when no formula is available.

Objective Probability Probability based on data.

Offer Price See Ask Price.

OIS See Overnight Indexed Swap.

OIS Rate The fixed rate exchanged for geometric average of overnight interest rates in an OIS swap.

Open-End Fund A mutual fund where the number of shares outstanding increases and decreases as investors deposit and withdraw funds.

Open Interest The total number of long positions outstanding in a futures contract (equals the total number of short positions).

Open Outcry A system of trading where traders meet on the floor of the exchange.

- Operational Risk** The risk of loss arising from inadequate or failed internal processes, people, and systems or from external events.
- Option** The right to buy or sell an asset.
- Organized Trading Facility** Electronic platform for standard over-the-counter derivatives in Europe.
- Originate-to-Distribute Model** Phrase used to describe the practice where a bank originates loans, credit card receivables, and so on, and securitizes them so that the credit risks are distributed to other investors.
- OTC Market** See Over-the-Counter Market.
- OTF** See Organized Trading Facility.
- Out-of-the-Money Option** Either (a) a call option where the asset price is less than the strike price or (b) a put option where the asset price is greater than the strike price.
- Outside-Model Hedging** Hedging against parameters that are assumed to be constant in the model (e.g., hedging against volatility in Black–Scholes–Merton).
- Overcollateralization** A situation where total principal of the underlying assets is greater than the sum of the principals of the tranches that are created from the assets.
- Overnight Indexed Swap** An agreement to exchange the geometric average of overnight rates during a certain period for a fixed rate.
- Over-the-Counter Market** A market where traders deal directly with each other rather than through an exchange. The traders are usually financial institutions, corporations, and fund managers.
- P2P Lending** Peer-to-peer lending is lending money to a borrower using an online platform without a bank or similar intermediary being involved.
- P&L Attribution** The analysis of the reason for daily profit and loss.
- Parallel Shift** A movement in the yield curve where each point on the curve changes by the same amount.
- Partial Duration** The percentage change in value of a portfolio for a small change in one point on the zero-coupon yield curve.
- Partial Expected Shortfall** The expected shortfall for part of a portfolio.
- Partial Simulation Approach** An approach for calculating value at risk where Greek letters and Taylor series expansions are used to approximate the change in the portfolio value.
- Par Value** The principal amount of a bond.
- Par Yield** The coupon on a bond that makes its price equal the principal.
- Payoff** The payment received by the holder of an option or other instrument.
- PD** Probability of default.
- Peak Exposure** A high percentile (e.g., 97.5) of the estimated exposure distribution at a particular future time.
- Performing Loan** A loan where interest is not more than 90 days overdue.
- Permanent Life Insurance** See Whole Life Insurance.
- Physical Default Probability** See Historical Default Probability.
- Plain Vanilla** A term used to describe a standard deal.

Poison Pill Action taken by a company to make a takeover of the company more difficult.

Poisson Distribution A distribution for number of events in a certain time period in a Poisson process.

Poisson Process A process describing a situation where events happen at random. The probability of an event in time Δt is $\lambda \Delta t$ where λ is the intensity of the process.

Policyholder The holder of an insurance policy.

Portfolio Immunization Making a portfolio relatively insensitive to interest rates.

Portfolio Insurance Entering into trades to ensure that the value of a portfolio will not fall below a certain level.

Positive Feedback Trading Trading where assets are sold after a price decline and bought after a price increase.

Positive Semidefinite A condition that must be satisfied by a variance–covariance matrix for it to be valid.

Power Law Law describing the tails of many probability distributions that are encountered in practice.

Predatory Trading Trading by someone who thinks another market participant will do a similar trade in the near future and that the similar trade, if it happens, will move the market. May or may not be legal, depending on the circumstances.

Premium The price of an option.

Prime Broker A bank that clears a hedge fund's trades, lends the hedge fund money, and provides other services.

Principal The par or face value of a debt instrument.

Principal Components Analysis An analysis aimed at finding a small number of factors that describe most of the variation in a large number of correlated variables. (Similar to factor analysis.)

Private Placement Selling a new issue of securities to a small number of large financial institutions without offering it to the general public.

Property–Casualty Insurance Property insurance provides a payout in the event of loss of property or damage to property. Casualty insurance provides protection against legal liabilities or damage to the property of others.

Proprietary Trading Trading for a financial institution's own account rather than the account of one of the financial institution's clients.

Public Offering Attempting to sell a new issue of securities to the general public.

Put–Call Parity The relationship between the price of a European call option and the price of a European put option when they have the same strike price and maturity date.

Put Option An option to sell an asset for a certain price by a certain date.

Puttable Bond A bond where the holder has the right to sell it back to the issuer at certain predetermined times for a predetermined price.

Puttable Swap A swap where one side has the right to terminate early.

QIS See Quantitative Impact Study.

Quadratic Model A quadratic relationship between change in portfolio value and percentage changes in market variables.

Quantitative Impact Study A study by the Basel Committee of the effect of proposed new regulations.

RAROC Risk-adjusted return on capital.

RCSA Risk control and self-assessment. An approach to assessing operational risk.

Real World The world we live in. Can be contrasted with the risk-neutral world, which is an artificial world used for valuing derivatives.

Rebalancing The process of adjusting a trading position periodically. Usually the purpose is to maintain delta neutrality.

Recovery Rate The amount recovered in the event of a default as a percent of the face value.

Reference Entity A company or country whose default is being insured against in a credit default swap.

RegTech The use of technology by a financial institution to improve compliance with regulations.

Regulatory Arbitrage Transactions designed to reduce the total regulatory capital of the financial institutions involved.

Regulatory Capital Capital a financial institution is required by regulators to keep.

Rehypothecation The use of collateral posted by one counterparty to satisfy the collateral requirements of another counterparty.

Reinsurance A situation where an insurance company passes on some of its risks to another company. Of course, it has to pay the other company for taking on the risks.

Reintermediation The development of new intermediaries following disintermediation.

Repo Repurchase agreement. A procedure for borrowing money by selling securities to a counterparty and agreeing to buy them back later at a slightly higher price.

Repo Rate The rate of interest in a repo transaction.

Reserve Requirement Percentage of deposits that must be kept in cash or on deposit with the central bank.

Reset Date The date in a swap or cap or floor when the floating rate for the next period is set.

Retail Banking Taking relatively small deposits from retail customers and making relatively small loans to them.

Reverse Stress Testing Using an algorithm to search for scenarios that lead to large losses.

Reversion Level The level to which the value of a market variable (e.g., a volatility or interest rate) tends to revert.

Reversion Rate The rate at which a market variable is pulled to the reversion level.

Rho The rate of change of the price of a derivative with the interest rate.

Right-Way Risk A situation where a counterparty becomes less likely to default as the dealer's exposure to the counterparty increases.

Risk Appetite A statement by a financial institution describing the level of risk that can be tolerated.

Risk Culture How risks are taken into account within an organization. An important aspect of a company's risk culture is the extent to which its employees focus on short-term profits without considering longer-term risks.

Risk Factor A term used to describe a market variable that changes through time, such as an exchange rate, commodity price, or interest rate.

Risk Weights In the weighted sensitivity approach, risk weights are a multiple of the daily standard deviation or daily volatility of a risk factor.

Risk-Free Rate The rate of interest that can be earned without assuming any risks.

Risk-Neutral Default Probability The default probability implied from a credit spread.

Risk-Neutral Valuation The valuation of an option or other derivative assuming the world is risk neutral. Risk-neutral valuation gives the correct price for a derivative in all worlds, not just in a risk-neutral world.

Risk-Neutral World A world where investors are assumed to require no extra return on average for bearing risks.

Risk-Weighted Amount See Risk-Weighted Assets.

Risk-Weighted Assets Quantity calculated in Basel I and Basel II. Total capital must be at least 8% of risk-weighted assets.

Robo-Advisor Online provider of wealth management services.

Robotics Process Automation A software application where a computer is trained to interact with existing computer systems.

Roll Back See Backward Induction.

RSF Factor Required stable funding factor, a weighting factor applied to funding requirements in the calculation of the Net Stable Funding Ratio.

SA-CCR A method for calculating exposure at default when standardized approaches are used. It is more elaborate than the current exposure method.

Sarbanes-Oxley An act passed in the United States in 2002 increasing the responsibilities of directors, CEOs, and CFOs of public companies.

Scenario Analysis An analysis of the effects of possible future scenarios on the value of a portfolio. Also used to estimate possible operational risk losses.

Scorecard Approach An assessment procedure used for operational risk.

SEC Securities and Exchange Commission.

Securitization A process whereby securities are created from future cash flow streams.

SEF See Swap Execution Facility.

Senior Tranche A tranche that is the last to experience losses because of defaults on the underlying assets.

Shadow Banking Bank services provided by non-banks.

Short Position A position assumed when traders sell shares they do not own.

Short Selling Selling in the market shares that have been borrowed from another investor.

SIFI Systemically important financial institution.

SIMM See Standard Initial Margin Model.

Simulation See Monte Carlo Simulation or Historical Simulation.

- Skewness** The measure of the lack of symmetry in a probability distribution.
- Solvency I** The current regulatory framework for insurance companies in the European Union.
- Solvency II** A new regulatory framework for insurance companies proposed by the European Union.
- Solvency Risk** The risk that liabilities will be greater than assets.
- Sovereign Wealth Funds** Funds set up to make investments for a country.
- Specific Risk Charge** Capital requirement for idiosyncratic risks in the trading book.
- Spectral Risk Measure** Risk measure that assigns weights to the quantiles of the loss distribution.
- Speculative Grade** See Non-Investment Grade.
- Speculator** An individual who is taking a position in the market. Usually the individual is betting that the price of an asset will go up or that the price of an asset will go down.
- Spot Interest Rate** See Zero-Coupon Interest Rate.
- Spot Price** The price for immediate delivery.
- Spot Volatilities** The volatilities used to price a cap when a different volatility is used for each caplet.
- SR 11-7** Guidance on model risk management provided by the Federal Reserve Board.
- Standard Initial Margin Model** A model developed by ISDA for calculating initial margin required by regulators on a portfolio of bilaterally cleared derivatives between two financial institutions.
- Standard Normal Distribution** A normal distribution with mean equal to zero and standard deviation equal to one.
- Standardized Measurement Approach** A new method for calculating operational risk capital proposed by the Basel Committee.
- Static Hedge** A hedge that does not have to be changed once it is initiated.
- Static Options Replication** A procedure for hedging a portfolio that involves finding another portfolio of approximately equal value on some boundary.
- Stochastic Variable** A variable whose future value is uncertain.
- Stock Index** An index monitoring the value of a portfolio of stocks.
- Stock Index Futures** Futures on a stock index.
- Stock Index Option** An option on a stock index.
- Stock Option** An option on a stock.
- Stop-Loss Rule** A trading rule where a position is unwound when losses on the position reach a certain level.
- Storage Costs** The costs of storing a commodity.
- Stressed ES** Expected shortfall calculated from a 250-day period of stressed market conditions that would be particularly bad for the financial institution.
- Stressed VaR** Value at risk calculated from a 250-day period of stressed market conditions that would be particularly bad for the financial institution.
- Stress Testing** Testing the impact of extreme but plausible market moves on the value of a portfolio.

Strike Price The price at which the asset may be bought or sold in an option contract. (Also called the exercise price.)

Structured Product A derivative designed by a financial institution to meet the needs of a client.

Student's t -Copula Copula based on the multivariate Student's t -distribution.

Student's t -Distribution Distribution with heavier tails than the normal distribution.

Subjective Probability A probability that reflects a person's opinion, rather than data.

Subprime Mortgage A mortgage that is more risky than the average mortgage.

Supplementary Capital See Tier 2 Capital.

Survivor Bond See Longevity Bond.

Swap An agreement to exchange cash flows in the future according to a prearranged formula.

Swap Execution Facility An electronic platform for standard over-the-counter derivatives in the United States.

Swap Rate The fixed rate in an interest rate swap that causes the swap to have a value of zero.

Swaption An option to enter into an interest rate swap where a specified fixed rate is exchanged for floating.

Swap Zero Curve See LIBOR/Swap Zero Curve.

Synthetic CDO A CDO where tranche payments depend on defaults in a bond portfolio, but the bond portfolio is not created.

Synthetic Option An option created by trading the underlying asset.

Systematic Risk Risk that cannot be diversified away.

Systemic Risk The risk that a default by one financial institution will lead to defaults by other financial institutions, creating risks for the whole financial system.

Tail Correlation A correlation between the tails of two distributions. Measures the extent to which extreme values tend to occur together.

Tail Dependence The measure of the frequency with which extreme outcomes occur for two variables at the same time.

Tail Loss See Expected Shortfall.

Taylor Series Expansion For a function of several variables, this relates changes in the value of the function to changes in the values of the variables when the changes are small.

Teaser Rate A low interest rate offered for the first two or three years on a mortgage.

Temporary Life Insurance See Term Life Insurance.

Terminal Value The value at maturity.

Term Life Insurance Life insurance that pays off if death occurs before a certain date.

Term Structure of Interest Rates The relationship between interest rates and their maturities.

Theta The rate of change of the price of an option or other derivative with the passage of time.

Threshold The value of outstanding transactions that leads to collateral being required in bilateral OTC trading.

- Tier 1 Equity Capital** Equity capital, redefined under Basel III (also called Tier 1 Core Capital).
- Tier 2 Capital** Subordinated debt (life greater than five years) and similar sources of capital.
- Tier 3 Capital** Short-term subordinated debt (life between two and five years).
- Time Decay** See Theta.
- Time Value** The value of an option arising from the time left to maturity (equals an option's price minus its intrinsic value).
- Total Expense Ratio** The ratio of total expenses to premiums for a property-casualty insurance company.
- Total Return Swap** A swap where the return on an asset such as a bond is exchanged for LIBOR plus a spread. The return on the asset includes income such as coupons and the change in value of the asset.
- Tracking Error** A measure of the error in a trading strategy designed to track a stock index.
- Trading Book** The part of a bank's portfolio that consists of instruments that are held for resale.
- Trading Days** Days when markets are open for trading.
- Tranche** One of several securities created from a portfolio. Examples are the tranches of an ABS.
- Transaction Costs** The cost of carrying out a trade (commissions plus the difference between the price obtained and the midpoint of the bid-offer spread).
- Treasury Bill** A short-term non-coupon-bearing instrument issued by the government to finance its debt.
- Treasury Bond** A long-term coupon-bearing instrument issued by the government to finance its debt.
- Treasury Note** A treasury bond lasting less than 10 years.
- Tree** A representation of the evolution of the value of a market variable for the purposes of valuing an option or other derivative.
- Uncleared Trades** Trades that are cleared bilaterally rather than through a central counterparty.
- Unconditional Default Probability** The default probability for a future time period as seen today. Can be contrasted with a conditional default probability, where the probability is conditional on no earlier default.
- Underlying Variable** A variable that the price of an option or other derivative depends on.
- Unit Trust** See Mutual Fund.
- Universal Life Insurance** A form of whole life insurance where the premium payments can be varied from year to year. The final payout is determined from the premiums, the performance of the investment, and costs.
- Unsystematic Risk** See Nonsystematic Risk.
- Up-and-In Option** An option that comes into existence when the price of the underlying asset increases to a prespecified level.

Up-and-Out Option An option that ceases to exist when the price of the underlying asset increases to a prespecified level.

Value at Risk A loss that will not be exceeded at some specified confidence level.

VaR See Value at Risk.

Variable Annuity An annuity (usually for life) where the amount received depends on the performance of a portfolio.

Variable Life Insurance A form of whole life insurance where the policyholder decides how funds will be invested. If the investments do well, the proceeds can be used to reduce premiums. There is a guaranteed minimum payout on death, but the payout may be higher if the investments do well.

Variance-Covariance Matrix A matrix showing variances of, and covariances between, a number of different market variables.

Variance Rate The square of volatility.

Variation Margin The extra margin required to bring balance in the margin account up to the required level.

Vasicek's Model A model of default correlation based on the Gaussian copula. (Other models developed by Vasicek are also sometimes referred to as Vasicek's model.)

Vega The rate of change in the price of an option or other derivative with volatility.

Vega-Neutral Portfolio A portfolio with a vega of zero.

VIX Index An index of the implied volatilities of options on the S&P 500.

Volatility A measure of the uncertainty of the return realized on an asset.

Volatility Scaling A procedure for scaling a historical daily change up or down to reflect differences between the current volatility and the volatility at the time of the historical daily change.

Volatility Skew A term used to describe the volatility smile when it is nonsymmetrical.

Volatility Smile The variation of implied volatility with strike price.

Volatility Surface A table showing the variation of implied volatilities with strike price and time to maturity.

Volatility Term Structure The variation of implied volatility with time to maturity.

Volcker Rule Part of the Dodd-Frank Act in the United States that prohibits proprietary trading.

Waterfall Rules for determining how principal and interest cash flows from an underlying portfolio are distributed to tranches. In a typical arrangement, interest cash flows are first used to pay the senior tranche its promised return. The cash flows (if any) that are left over are used to provide the next-most-senior tranche with its promised return, and so on. Principal cash flows are usually used first to repay the senior tranche, then the next-most-senior tranche, and so on. The equity tranche is the last to receive both interest and principal cash flows.

Weather Derivative A derivative where the payoff depends on the weather.

Weighted Sensitivity Terminology used for delta in the weighted sensitivity approach.

Whole Life Insurance Life insurance that lasts for the whole life of the insured and so is certain to provide a payout.

Wholesale Banking Taking large deposits and making large loans.

Within-Model Hedging Hedging only against those variables that the model considers as stochastic (cf. Outside-Model Hedging).

Worst Case Default Rate (WCDR) The default rate that will not be exceeded at some confidence level.

Writing an Option Selling an option.

Wrong-Way Risk A situation where a counterparty becomes more likely to default as the dealer's exposure to the counterparty increases.

Yield A return provided by an instrument.

Yield Curve See Term Structure of Interest Rates.

Zero-Coupon Bond A bond that provides no coupons.

Zero-Coupon Interest Rate The interest rate that would be earned on a bond that provides no coupons.

Zero-Coupon Yield Curve A plot of the zero-coupon interest rate against time to maturity.

Zero Curve See Zero-Coupon Yield Curve.

Zero Rate See Zero-Coupon Interest Rate.

Z-Score A number indicating how likely a company is to default.

RMFI Software

The RMFI software version 1.00 that accompanies this book can be downloaded from the author's website: www-2.rotman.utoronto.ca/~hull/software.

Getting Started

The most difficult part of using software is getting started. Here is a step-by-step guide to valuing an option using the RMFI software.

1. Open the Excel file RMFI Software v1.00 that you have downloaded.
2. If you are using Office 2007, click on *Options* at the top of your screen (above the F column) and then click *Enable content*. If you are not using Office 2007, make sure that your security for Macros is set at medium or low. (You can do this by clicking *Tools*, followed by *Macros*, followed by *Security*.)
3. Click on the *Equity_FX_Indx_Fut_Opts_Calc* worksheet tab at the bottom of the page.
4. Choose *Currency* as the Underlying Type and *Binomial: American* as the Option Type. Click on the *Put* button. Leave *Implied Volatility* unchecked.
5. You are now all set to value an American put option on a currency. There are seven inputs: exchange rate, volatility, risk-free rate (domestic), foreign risk-free rate, time to expiration (yrs), exercise price, and time steps. Input these in cells D5, D6, D7, D8, D13, D14, and D15 as 1.61, 12%, 8%, 9%, 1.0, 1.60, and 4, respectively.
6. Hit *Enter* on your keyboard and click on *Calculate*. You will see the price of the option in cell D20 as 0.07099 and the Greek letters in cells D21 to D25.
7. Click on *Display Tree*. You will see the binomial tree used to calculate the option.

Next Steps

You should now have no difficulty valuing other types of options on other underlyings with this worksheet. To imply a volatility, check the *Imply Volatility* box and input the option price in cell D20. Hit *Enter* and click on *Calculate*. The implied volatility is displayed in cell D6.

Many different charts can be displayed. To display a chart, you must first choose the variable you require on the vertical axis, the variable you require on the horizontal axis, and the range of values to be considered on the horizontal axis. Following that, you should hit *Enter* on your keyboard and click on *Draw Graph*.

Other points to note about this worksheet are:

1. For European and American equity options, up to 10 dividends on the underlying stock can be input in a table that pops up. Enter the time of each dividend (measured in years from today) in the first column and the amount of the dividend in the second column.
2. Up to 500 time steps can be used for the valuation of American options, but only a maximum of 10 time steps can be displayed.
3. Greek letters for all options other than standard calls and puts are calculated by perturbing the inputs, not by using analytic formulas.
4. For an Asian option, the *Current Average* is the average price since inception. For a new deal (with zero time to inception) the current average is irrelevant.
5. In the case of a lookback option, *Minimum to Date* is used when a call is valued and *Maximum to Date* is used when a put is valued. For a new deal these should be set equal to the current price of the underlying asset.
6. Interest rates are continuously compounded.

CDSs

The CDS worksheet is used to calculate hazard rates from CDS spreads and vice versa. Users must input a term structure of interest rates (continuously compounded) and either a term structure of CDS spreads or a term structure of hazard rates. (All data for hazard rates, CDS spreads, and interest rates must be entered in chronological order.) The initial hazard rate applies from time zero to the time specified for the first hazard rate; the second hazard rate applies from the time corresponding to the first hazard rate to the time corresponding to the second hazard rate; and so on. The calculations are carried out assuming that default can occur only at points midway between payment dates. This corresponds to the calculations for the example in Appendix K.

CDOs

The CDO worksheet calculates quotes for the tranches of CDOs from tranche correlations input by the user (see Appendix L). The attachment points and detachment points

for tranches are input by the user. The quotes can be in basis points or involve an upfront payment. In the latter case, the spread in basis points is fixed and the upfront payment, as a percent of the tranche principal, is either input or implied. (For example, the fixed spread for the equity tranche of iTraxx Europe or CDX NA IG is 500 basis points.) The number of integration points defines the accuracy of calculations and can be left as 10 for most purposes. (The maximum is 30.) Data for interest rates and hazard rates must be entered in chronological order. The software displays the expected loss as a percent of the tranche principal (ExpLoss) and the present value of expected payments (PVPmts) at the rate of 10,000 basis points per year. The spread is $\text{ExpLoss} \times 10,000 / \text{PVPmts}$. The upfront payment is $\text{ExpLoss} - (\text{Spread} \times \text{PVPmts} / 10,000)$. The worksheet can be used to imply either tranche (compound) correlations or base correlations from quotes input by the user. For base correlations to be calculated, it is necessary for the first attachment point to be 0% and the detachment point for one tranche to be the attachment point for the next tranche.

Equity Option Functions

The equity option functions worksheet provides functions for valuing a range of different options. The functions can be used in conjunction with the user's own worksheets to allow the properties of the options to be investigated.

Value at Risk

The value at risk worksheet allows VaR to be calculated for a portfolio of options dependent on a single stock using the linear model, a model based on the first three moments, and an analytic approach where equally likely values for a portfolio are generated. The first two approaches are based on the material in Chapter 14.

Definition of Greek Letters

The Greek letters are defined as follows.

Delta: Change in option price per dollar increase in underlying asset.

Gamma: Change in delta per dollar increase in underlying asset.

Rho: Change in option price per 1% increase in interest rate (e.g., interest increases from 5% to 6%).

Theta: Change in option price per calendar day passing.

Vega: Change in option price per 1% increase in volatility (e.g., volatility increases from 20% to 21%).

Table for $N(x)$ When $x \geq 0$

This table shows values of $N(x)$ for $x \geq 0$. The table should be used with interpolation. For example, $N(-0.6278) = N(0.62) + 0.78[N(0.63) - N(0.62)] = 0.7324 + 0.78 \times (0.7357 - 0.7324) = 0.7350$.

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

Table for $N(x)$ When $x \leq 0$

This table shows values of $N(x)$ for $x \leq 0$. The table should be used with interpolation. For example, $N(-0.1234) = N(-0.12) - 0.34[N(-0.12) - N(-0.13)] = 0.4522 - 0.34 \times (0.4522 - 0.4483) = 0.4509$.

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233

Index

- 5/30 swap, example, 121–122
- 1996 Amendment (Basel Committee), 542–545
- ABACUS (Goldman Sachs), issues, 522–523
- Abu Dhabi Investment Authority, 43
- Accounting, 40, 492
 - ratios, 368
- Accrual payment, present value (calculation), 690t
- Accumulation value, 52
- Actively traded products
 - model risk sources, 488–489
 - pricing models, 487–490
- Active management, passive management
 - (contrast), 80–81
- Additional Tier 1 (AT1) capital, 568
- Add-on factor, 537
- Adjoint differentiation, usage, 401
- Adjustable rate mortgages (ARMs), 147
- Adoboli, Kweku, 123, 641, 643
- Advanced IRB approach (Basel II), 546, 549–551
- Advanced measurement approach (AMA),
 - 431, 555
- Adverse market, risk, 456
- Adverse selection, 62, 445
- Advisory services, 34
 - aspects, 36
- Agency costs, 157
- Akepa, greenwashing, 510
- Alipay, services, 624
- Allied Irish Bank, losses, 440, 641, 642
- Alpha, 12–13
 - multiplier, 402
 - portfolio manager creation, 13
 - positive alpha, 81
 - representation, 13
- Alternative investments, 84
- Alternative risk-return combinations, 5f
- Alternative risky investments, 4f
- Alternative ticke rule, 101–102
- Altman's Z-score, 368–369
- Amaranth (hedge fund), 87
- Amazon One, 625
- Ambac, downgrade triggers (impact), 398
- AMERIBOR, 295
- American Depository Receipts (ADRs), 92
- American International Group (AIG)
 - collateral posting, absence, 398
 - downgrade triggers, impact, 398
 - failure, 373, 402
 - losses, 66, 137
 - OTC market losses, 137
- American options, 112
 - valuation, 673
- American put, binomial tree, 674f
- AMP Capital Investors, 148
- Anchoring and adjustment, 524
- Annual compounding, usage, 307
- Annual percentage default rate, 208t

- Annual renewable term, 48
- Annuity
 - contracts, 51–52, 63
 - profitability, 56
 - payments, 52
 - products, sale, 53
- Apple Pay, services, 624
- Application program interfaces (APIs), 628–629
- Arbitrage, 90–2
 - regulatory arbitrage, 157
- Arbitrage pricing theory (APT), 13–14
- Archegos Capital Management, losses, 642
- Arithmetic mean, 82, 707
- Ashanti Goldfields, hedging problems, 462
- Asian basket option, 121
- Asian monetary crisis (1997–1998), 471
- Asian options, 119, 337
- Ask prices, function, 451f
- Ask quotes, relationship, 450, 451f
- Asset-backed commercial paper (ABCP), liquidity
 - backstop arrangements, 460–461
- Asset-backed securities (ABSs), 145, 150–152
 - creation, 150f, 152f
 - losses, 153f, 155–156
 - mezzanine tranche, rating issue, 155
 - waterfall, 151f
- Asset-backed security collateralized debt obligation
 - (ABS CDO), 152–153, 373, 522
 - creation, 152f
 - losses, 153f, 155–156
 - mezzanine tranches, 153
 - risk characteristics, 155
 - usage, 153–155
- Asset-liability management group, role, 304
- Assets
 - forward price, volatility, 406
 - long/short positions, 99–102
 - maturities, rates (increase), 304t
 - normal probability, translation, 339f, 340f
 - portfolio investment, 283–284
 - price
 - daily volatility, asset price calculation, 164
 - volatility, relationship, 218–219
 - proportional bid-ask spread, 453
 - return, variance (calculation), 166
 - sale price, factors, 450
 - securitization, ability, 458
 - shorting, 322
 - swaps, structuring, 379
 - volatility, 386
- Assets under management, 76
- Asymmetric information problem, 376
- AT&T
 - return distribution, 287
 - shares portfolio, 281
- At-the-money options, 112
- Autocorrelations
 - exhibition, 184
 - GARCH model, impact, 184t
 - impact, 244–245
- Auto insurance premium, increase, 61
- Automated teller machines (ATMs),
 - offering, 39
- Autoregressive conditional heteroskedasticity (ARCH), 163, 175
- Availability, 525
- Available stable funding (ASF) factor,
 - 463, 572t
- Average cumulative issuer-weighted default rates, 370t
- A.W. Jones & Co., 84
- Back-end load, 78
- Back office, separation, 645
- Back-testing, 93, 249–252, 482, 544–545, 594
- Balance sheet
 - abbreviated balance sheet, 63t
 - alternative, 29t
 - ratios, 368
 - summary, 27t
- Bankers Trust (BT), 121–122, 520, 646–647
- Bank for International Settlements (BIS), 102
 - Accord (1988), 536–539
 - CDS statistics, 373
- Bank Holding Company Act, Douglas
 - Amendment, 27
- Banking
 - book, 40
 - trading book, contrast, 595–596
 - commercial banking, 26–28
 - conflicts of interest, 38–39
 - investment banking, 31–36
 - open banking, 628–629
- Bank of England, Federal Reserve Board, 461
- Bankruptcy
 - CCPs, relationship, 141–142
 - chance, increase, 38
 - costs, 15–17
 - hidden costs, 15–16
 - regulation, 17

- Banks, 25
 - assets
 - components, 537, 548
 - loans components, 553
 - balance sheet, 573
 - business, lending (emphasis), 601
 - climate change, impact, 502–504
 - climate risks, impact, 503t
 - commercial bank, capital requirements, 28–30
 - conflict-of-interest rules, violations, 39
 - diversification, 19
 - estimated loss, 437
 - exposures, 551–553
 - fees, reduction, 624
 - fixed-rate loans, usage, 111
 - floating-rate deposits, usage, 111
 - forward contract purchase, 406
 - funding strategy, 464–465
 - interest rate swap, 538
 - loans, default probabilities, 240
 - pre-1988 bank regulation, 535–536
 - regulation, 26
 - reasons, 533–535
 - residential mortgages assets, 554
 - risks, 42–43
 - status, 39–42
 - stress testing
 - methodologies, enhancement, 355
 - recommendations, 354–356
 - trading book, 542
- Barings
 - default, 534
 - losses, 641, 642
- Barrier options, 119
- Basel Accord. *See* BIS Accord
- Basel Committee
 - 1996 Amendment, 542–545
 - formation, 535–536
 - grid approach, 138, 697–698
- Basel I, 533
- Basel II, 533, 545–546
 - advanced IRB approach, 546, 549–551
 - advanced measurement approach, 555
 - basic indicator approach, 555
 - credit risk capital, 546–555
 - Foundation Internal Ratings Based (IRB)
 - approach, 546, 549–551
 - market discipline, 556–557
 - operational risk, 555–556
 - pillars, 545
 - standardized approach, 546–548, 554, 555
 - risk weights, 547t
 - supervisory review, 556
- Basel II.5, 563, 564–567
- Basel III, 539, 563, 567–575
- Basic indicator approach, 431
- Basic Indicator Approach (Basel II), 555
- Basis points, measure, 306
- Basket options, 119
- Bayesian networks, 360
- Bayes' theorem, 360, 525
- Bear Stearns, problems, 305
- Berkshire Hathaway, 62
 - climate risk metrics, disclosure (reluctance), 507
- Best efforts, 32
- Best-efforts deal, 32
- Best-fit value, 179
- Beta, 9–10
 - estimation/calculation, 272, 516
 - gamma, relationship, 699t
 - valuation, 12
 - values, search, 274
 - zero value, 10, 89
- Beta-neutral fund, 89
- Bid-ask spread, 450
 - daily trading volumen function, 456
 - high level, 452
 - mean/standard deviation, 454
- Bid-offer spread, 111, 394
- Bid prices, 130
 - function, 451f
- Bid quotes, relationship, 450, 451f
- Bilateral cleaning, 132f, 134–135
- Binary options, 119, 220
 - risk-neutral valuation, application, 221–222
- BINOMDIST, 251
- BIS Accord (Basel Accord) (1988), 536, 540–541
- BIS Amendment, 234
- Bitcoin (BTC), value (2015–2022), 626f
- Bivariate normal distribution, 206
 - random samples, 206f
- Bivariate Student's *t*-distribution, 205
- Bivariate Student's *t*-distribution, random
 - samples, 207f
- Black, Fischer, 632
- BlackRock
 - assets under management, 76
 - net zero objective, 507
- Black-Scholes-Merton formulas, 222, 326, 386, 669

- Black-Scholes-Merton model, 487
 - application, 222
 - in-the-money prediction, 169
 - reality description, 488
- Black-Scholes-Merton option pricing model, 165
- Blockchain, 622–624
- Bloomberg, Michael, 500
- Bloomberg Short-Term Yield Index (BSBY), 295
- BNP Paribas, fines, 431, 635
- Boesky, Ivan, 91
- Bogle, John, 631
- Bogle's folly, 77
- Bonds
 - average credit spread, 383
 - basis, 379–380
 - bond-yield spread, 380
 - buy and hold strategy, 425
 - constant level of risk strategy, 425
 - default, loss calculation, 381t
 - duration/convexities, 309f
 - funds, 77
 - issue, arrangement, 38
 - market crash (1994), 471
 - price, 309
 - pricing, 661
 - value, 309
 - yields, 68, 661
 - CDS spreads, relationship, 378
- Book, knowledge, 319
- Bootstrap method, 269–270, 380–381, 662
- Borrowers, online platforms (providing), 629
- Borrowing, ability, 458, 460
- Bottom-up approach, 601
- Brainstorming, 350, 353–354
- Break-even annual premium, 56
- Break-even premium, 55–56
 - payment, 56
- Breaking the buck, 76
- Breakout trading, 467
- Brent crude oil, 117
- Brin, Sergey, 35
- British pounds, sale (example), 99
- Broken arrow stress test, 350
- Brokers
 - assistance, 37
 - commissions, payment, 59
 - investor instruction, 100
- Brundtland, Gro Harlem, 508
- Bucket deltas, 314
- Bucketing, usage, 315f
- Buffett, Warren, 62, 84, 507
- Bunching, 252
- Business days, calendar days (contrast), 165
- Business disruption, 434
- Business Indicator Component (BIC), 439
 - calculation, 439t
- Business practices, 434
- Business risk economic capital, 603
- Buy and hold strategy, 425
- Buying on margin, 100
- CAC 40, 259, 267, 268, 284
- Calendar days, business days (contrast), 165
- Calibrating instruments, selection, 490
- California earthquake, losses, 58
- Call option, 112
 - delta, 324f
 - expected payoff, calculation, 219
 - value, 323f
- Call, purchase/sale, 114
- Cap and trade, 501
- Capital
 - adequacy, 29–30, 558
 - conservation buffer, 567, 569–570
 - dividend restrictions, 569t, 570t
 - definition/requirements, 567, 568–569
 - markets, flight to quality, 352
 - regulatory requirements, 601f
 - requirements, 63–64, 538–539
 - sources, 30
 - structure, aggressiveness, 30
- Capital asset pricing model (CAPM), 8–13, 80
 - assumptions, 10–11
 - beta, 247
 - example, 9f
 - usage, 226
- Capital gains, taxation, 11
- Capital Requirement Directive 2 (CRD2), 580
- Capital value adjustment (KVA), 394, 408
- Capitol riots (2021), 37
- Carbon dioxide emissions, reduction, 501
- Carbon pricing, strengthening, 506
- Carney, Mark, 500, 607
- Cascade approach, 593
- Cash
 - holdings, 458
 - management, 88
- Cash flows
 - example, 101t, 110t
 - processing, 629

- Casualty insurance, protections, 57
- Catastrophe (CAT) bonds, 58–59
- Catastrophic risk, 58
 - scenarios, 558
- Causal relationships, 441
- Cayman Islands, tax-favorable jurisdictions, 84–85
- CDX NA IG portfolio, 376–377
- Central bank, borrowing, 458, 461
- Central bank digital currencies (CBDCs), 628
- Central clearing, 132f, 133–134
- Central clearing parties (CCPs), 393, 578–579
- Central counterparty (CCP), 131
 - bankruptcy, relationship, 141–142
 - equity, 134
- Charter provisions, addition, 36
- Cheapest-to-deliver bond, 376
 - option, 379
- Chicago Board Options Exchange (CBOE)
 - stock option trading, 112–114
 - trading focus, 130
- Chicago Mercantile Exchange
 - live cattle futures contracts, trading, 105
 - weather futures trading, 116–117
- Chi-square distribution, 252
- Cholesky decomposition, 201
- Citigroup, losses, 156
- Citron, Robert, 641, 643, 644, 650
- Claims, payouts, 65
- Clawback clause, 85
- Clearing house
 - initial/variation margin requirement, 108
 - members, requirement, 131
- ClearPort (CME), 133, 141
- Clients, product sales (appropriateness), 646–647
- Climate awareness bond, issuance, 506
- Climate risk, 497–507
 - disclosures, introduction (commitment), 507
- Climate warming, control (policies), 506
- Closed-end funds, 79
- CME Group, 117
- Cognitive biases, 524–526
- Coherent risk measures, 238–242
- Cohn, Gary, 522
- Co-insurance provision, 62, 444
- Collateral, 131, 395
 - adjustments, 548
 - calculation, 548
 - cure periods, 396–397
 - posting, 88, 135
 - risk weight adjustments, simple
 - approach/comprehensive approach, 548
- Collateralized debt obligations (CDOs), usage, 152–155
- Combined ratio after dividends, 59
- Combined ratio, property-casualty insurer
 - calculation, 59
- Commercial bank, capital requirements, 28–30
- Commercial banking, 26–28
 - US ranking, 26t
- Commercial paper (CP) rate, payment, 121–122
- Committee of Sponsoring Organizations of the Treadway Commission (COS), ERM Integrated Framework, 513–514
- Commodity
 - forward price, 664
 - price, consideration, 227
- Commodity Futures Trading Commission (CFTC), monitoring responsibility, 579
- Companies
 - annual percentage default rate, 208t
 - assets/equities, valuation, 386
 - equity, value/volatility, 387
 - insurance companies, role, 47
 - money lending, 38
 - risk, return (contrast), 14–17
 - share, recommendation, 38
 - sustainable/ethical brands, creation, 510
- Competitor, mode weaknesses (exploitation), 486
- Compliance, 634–636
- Component ES, 248
- Component measures, 247–248
- Component value at risk, 247
- Compounding frequency, 163
 - effect, 656t
- Compound options, 119
- Comprehensive approach (collateral risk weights adjustment), 548
- Comprehensive Capital Analysis and Review (CCAR), 356
- Comprehensive risk measure (CRM), 566–567
- Conceptual soundness, evaluation, 482
- Conditional value at risk (CVaR), 238
- Confidence level, 243, 246–247, 456, 600f
 - calculation, 273
 - impact, 237, 265
 - increase, 257
- Constant level of risk
 - assumption, 566
 - strategy, 425

- Constant Maturity Treasury (CMT) yield, 121–122
- Constant variance, estimation, 179–180
- Contagion, 385
- Contingent convertible bonds (CoCos), 576
- Continuous compounding, 218, 656–657
 - usage, 307
- Contracts
 - flaws, 375–376
 - purchase (Intel), net profit, 114f
- Contributory policy, 51
- Control areas, 118
- Controlling interest, obtaining, 36
- Convergence arbitrage, 648
- Convertible arbitrage, 91
- Convertible bonds, 31
- Convexity, 308–310
 - adjustment, 485, 486
 - bonds, convexities, 309f
 - dollar convexity, 310
- Cooke ratio, 536–539
- Cooling degree days (CDD), 116–117
- Copulas, 193, 202–207, 608
 - application, 202–203
 - approach, algebraic expression, 204–205
 - correlation, 209
 - joint distribution definition method, 205f
 - usage, 607–608
- Core variables, peripheral variables (contrast), 350
- Cornish-Fisher expansions, 341–342
- Corporate bonds
 - defaults, 64
 - recovery rates, 372t
- Corporate exposures, 551–553
- Corporate failures, 17
- CorporateMetrics, 234
- Corporate social responsibility (CSR), 505
- Correlation
 - coefficient, 201, 283
 - covariance matrices, relationship, 198–199
 - definition, 193–195
 - dependence, contrast, 194–195
 - estimation, 197
 - matrix, 198t, 286t
 - calculation, 285t
 - model, 421–423, 422f
 - monitoring, 195–197
- Correlation-dependent instruments, standardized
 - capital charge, 567t
- Cost-benefit analysis, 441
- Countercyclical buffer, 568, 570, 570t
- Counterparty
 - credit risk, 568, 574
 - default, 393–394
 - risk, 379
 - exposure (securing), collateral (usage), 548–549
- Coupon-bearing bond, duration, 306
- Covariance, 194
 - consistency condition, 198–199
 - matrix, calculation, 285t, 286t
 - rate, 195–196
- COVID-19 economic recovery plans,
 - ensuring, 507
- COVID-19 pandemic, impact, 167, 183, 503
- Credit contagion, 385
- Credit correlation, 424
 - dependence, 564
- Credit crisis (2007), 349, 460
- Credit crunch, 643
- Credit default swaps (CDSs), 372–377, 374f
 - bond basis, 379–380
 - contract
 - flaws, 375–376
 - restructuring clause, 380
 - counterparty default risk, 379
 - market, 373–374
 - marking to market, 691
 - payoff, 380
 - spreads, 694
 - bond yields, relationship, 378
 - valuation, 689
- Credit derivatives, 555
- Credit event, 372
- Credit exposures, 540
- Credit indices, 376–377
- CreditMetrics, 420–423
 - correlation model, 422f
 - Monte Carlo simulation trial, 421
- Credit ratings, 19, 367–369
 - agencies, 19
- Credit risk, 109, 595
 - capital (Basel II), 546–555
 - defining, 42
 - economic capital, 603
 - estimates, usage, 386
 - exposure, 536
 - handling, 134
 - loss probability density function, 604f
 - mitigation techniques, costs/benefits, 540
 - plus, 417–419
 - transfer, 156

- Credit Risk Plus model, 423, 425, 426, 603
- Credit risk VaR, time horizon, 414
- Credit spreads, 377–380
 - cumulative default probabilities, comparison, 382t
 - differences, 384–386
 - risk, 423–425, 595
 - term structure, 401
 - usage, 380–382
- Credit Suisse
 - CoCo bond issues, 576–577
 - VaR calculation methodology, 417
- Credit support annex (CSA), 134–135, 396
- Credit transition matrices, manipulation, 687–688
- Credit value adjustment (CVA), 394–399, 574
 - calculation, 397–398
 - risk, 400–401
- Credit value at risk (CVaR), 413
- Cross-border litigation, 581
- Cross gamma, 340
- Cross-selling, 517–519
- Crowded exit, 466
- Crowdfunding, 630–631
- Cryptocurrencies, 626–627
- Cumulative bivariate normal distribution, 200
- Cumulative default probabilities, credit spreads (comparison), 382t
- Cumulative joint probability
 - distribution, 204t
- Cumulative loss distribution, 236f
- Cumulative risk-neutral probability, 404
- Cure periods, 396–397
- Currency swaps, 666–667
 - interest rate swaps, contrast, 405
 - valuation, 667t
- Current ES, 263, 265
- Current exposure method (CEM), 537
- Current VaR, 263, 265
- Curry, Thomas J., 432
- Curvature risk charge, 589–590
 - calculation, 590f
- Daily exchange rate, absolute size
 - elevation, 168t
 - values, calculation, 171t
- Daily returns, correlation matrix (calculation), 285t
- Daily volatility
 - estimate, 266–267
 - monitoring, 172–175
- Danish Financial Supervisory Authority (DFSA),
 - traffic light solvency stress-testing system, 357–358
- Debit value adjustment (DVA), 395, 402–403
- Debt
 - issuance, 92
 - repudiation, 41
 - rescheduling, 41
- Dedicated shorts, 89–90
- Deductible, presence, 62, 444
- Deep-in-the-money, 336
- Deep-out-of-the-money, 336
- Deep-out-of-the-money call/put options,
 - purchase, 169
- Default
 - chance, 240
 - distance, 388
 - events, 135–136
 - intensity, 370
 - losses, probability distribution (shape), 420f
 - probability distribution, 419t
 - risk-neutral probability, 395
 - survival, 534–535
- Default fund contributions, 134
- Default probabilities
 - calculation, approximation, 380
 - credit risk estimates, usage, 386
 - estimates, comparison, 382–386
 - estimation, 367
 - credit spreads, usage, 380–382
 - equity prices, usage, 386–389
 - internal-ratings-based approaches, 368
 - historical default probabilities, 369–371
 - model, performance, 388
- Default rate (DR)
 - observations, 211
 - recovery rates, dependence, 372
 - sampling, 418
 - variation, 208
- Default risk charge, 591
- Deferred annuity, 52
- Defined benefit plans, 66
 - contributions, 69
 - viability, 68–69
- Defined contribution plans, 67
- Deleveraging, 469–470, 470f
- Delta (Δ), 297, 319–327
 - calculation, 314–315, 400, 675
 - exposures, 299t
 - hedging, 320, 326–327, 334, 337, 490

- Delta (Δ) (*continued*)
 - simulation, 325t, 326t
 - limit, 336
 - neutrality, 329, 337
 - risk, 589
 - sensitivity, 327–328
- Delta/gamma approximation, usage, 401
- Delta-neutral portfolio, 332
 - alternative relationships, 334f
 - gamma, 335
- Delta One team (UBS), 122–123
- Dependence
 - correlation, contrast, 194–195
 - possibilities, 195f
- Deposit insurance, 30–31
- Deposits, rolling over, 305
- Derivatives
 - credit exposure, 394–395
 - cumulative risk-neutral probability, 404
 - EAD calculation, 551
 - electricity derivatives, 118–119
 - exposures, 570
 - hedging, realities, 335
 - markets, 102–103
 - natural gas derivatives, 117–118
 - non-traditional derivatives, 115–119
 - oil derivatives, 117
 - OTC derivatives market, 129
 - plain vanilla markets, 103–115
 - portfolio, 542t
 - positions, 540
 - price change, rate (ρ) estimation, 211–212, 332
 - principal, add-on factors, 538t
 - risk, 319
 - P&G example, 122–123
 - single derivatives transaction, positive value, 403–404
 - situations, 394
 - transaction, impact, 399–400
 - value, 221–223
 - weather derivatives, 116–117
- Deutsche Bank
 - economic capital, 611, 612t
 - regulatory capital, 612t
- Digital Reasoning, surveillance software, 635–636
- Digital World Acquisition Corp (DWAC), SPACactions, 37
- Dimon, Jamie, 627
- Directed brokerage, 83
- Discrete outcomes, 222–223
- Disintermediation, 619
- Distance to default, 388
- Distressed securities, 90
- Distributed ledger technologies (DLTs), 624
- Distribution tail, estimation, 272
- Diversification
 - benefits, 282, 609
 - estimation, 644–645
 - improvement, 75–76
- Diversity, importance, 472
- Dodd-Frank Act, 37, 99, 156, 578–581
- Dodd-Frank Act Stress Test (DFAST), 356
- Dollar convexity, 310
- Dollar duration, 308
- Dollar-neutral fund, 89
- Domestic systemically important banks (D-SIBs), 356, 574–575
- Douglas Amendment, 27
- Dow Jones Industrial Average (DJIA), crash (1987), 468–469
- Downgrade triggers, 398–399
- Drexel default, 534
- Duration, 305–308
 - bonds, duration, 309f
 - calculation, 306t
 - definitions, 305–306
 - measures, 305
 - relationship, 307
- Dutch auction approach, 33–35
- DWS (Deutsche Bank), greenwashing, 510
- Dynamic hedging, 323, 336, 467
- Eastman Kodak, innovation (absence), 637
- Economic capital, 42–43, 599
 - aggregating, 606–609
 - allocation, 610–611, 612t
 - calculation, 600f
 - components, 602–603
 - definition, 600–601
 - estimates, 608t
 - measurement, 601
 - sum, 607
- Economic sustainability, 509
- Economies of scale, 51, 327
- Economy (stress), liquidity (impact), 457
- Effective Expected Exposure (Effective EE), 551
- Efficient frontier, 6–8
 - obtaining, 6f
- Eigenvectors/eigenvalues, 681–683
- Electricity derivatives, 118–119

- Electronic trading, 130
- E-mails/phone calls, impact, 442
- Emerging markets, 92
- Empirical distribution, approximation, 264
- Employees
 - compensation, 157
 - retirement, 67
 - stock options, granting, 36
- Employment practices/workplace safety, 434
- End-of-year bonuses, 157
- Endowment life insurance, 51
- Enron
 - bankruptcy, 445
 - counterparties, losses, 642
 - downgrade triggers, impact, 398
- Enterprise risk management (ERM), 513–514
- Environmental, Social, and Governance (ESG), 497, 507–508
- Environmental sustainability, 509
- Equal weighting approach, volatilities, 286t
- Equitable Life, annuity products sale, 53
- Equity
 - capital, worst-case erosion, 517
 - delta, 387
 - financing, 33
 - funds, 77
 - index, 52, 77
 - increase/decrease, 348
 - instantaneous volatility, 386
 - long/short equity, 89
 - prices, usage, 386–389
 - tranche, 150–151
- Equity-market-neutral fund, 89
- E*Trade, 37
- Euler's theorem, 248–249
- European call option, 119
 - consideration, 671
 - positions, 352–353
 - theta, variation, 331f, 332f
- European Central Bank (ECB)
 - supervisory climate stress test, 502
- European Central Bank (ECB), haircuts usage, 461
- European Growth Trust (EGT) fund, 606
- European options, 112
 - valuation, 353, 669
- European stock options, Black-Scholes-Merton
 - formulas, 222
- European Union, stress tests, 356
- Europe, insurance company regulation, 66
- Euro short-term rate (ESTER), 109, 294
- Euro Stoxx 50, 122
- Exception, probability, 250
- Excess kurtosis, 11
- Excess-of-loss reinsurance contract, 63
- Exchange rate
 - increase/decrease, 348
 - standard deviation (movement), probability (log-log plot), 171f
- Exchange-traded derivatives market, 102
- Exchange-traded funds (ETFs), 75, 79–80
 - advantages, 80
 - creation, 707
- Exchange-traded markets, 115, 129–131
 - OTC derivatives markets, convergence, 141
 - transactions, 134
- Exercise price, 114
- Exotic options, 119–120
 - hedging, 336–338
 - portfolio, 338
- Expected bankruptcy costs, 16
- Expected cash flows, 50f
- Expected cash outflow, 49
- Expected (average) future return
 - assumption, 219
 - estimation, 220
- Expected payments, present value (calculation), 690t
- Expected payoff
 - calculation, 220
 - discounting, 219, 220
- Expected payoff, present value (calculation), 690t
- Expected return, 2, 5f
 - calculation, 10
 - investor attention, 10–11
 - market return, relationship, 12, 12f
 - risk, trade-off, 2–3
- Expected shortfall (ES), 233, 238
 - aggregating, 249
 - calculation, 249, 262, 273, 302
 - historical simulation, usage, 262
 - parameters, selection, 242–247
 - procedure, extensions, 286–287
 - stressed ES, 263
- Expense ratio, 59, 77
- Expiration date, 112
- Exponentially weighted moving average (EWMA)
 - model, 175–177, 267
 - approaches, volatilities, 286t
 - estimation, 180–184

- Exponentially weighted moving average (EWMA) model (*continued*)
- selection, 179
 - usage, 196–197, 267t, 285–286
- Exponential spectral risk measure, 242
- Exposure at default (EAD), calculation, 417, 551–552, 577
- External credit ratings, usage, 579
- External data, internal data (contrast), 436–437
- External fraud, 434
- External historical data, inflation adjustment, 437
- Extreme value theory (EVT), 257, 270–273, 348
- applications, 273–276
 - calculations, 274t
 - result, 271
- Exxon *Valdez* oil spill (1997), 509
- Factor
- importance, 300
 - scores, 301
- Factor copula model, 207
- Factor models, 201
- Factor neutrality, 89
- Factor scores, standard deviation, 300t
- Fear index, 167
- Features, 620
- Federal Deposit Insurance Corporation (FDIC), 65–66
- creation, 30
 - liquidation powers, 578
- Federal Home Loan Mortgage Corporation (FHLMC) (Freddie Mac), 41, 461
- Federal Insurance Office (FIO), creation/setup, 66, 578
- Federal National Mortgage Association (FNMA) (Fannie Mae), 41, 461
- Federal Reserve
- annual stress tests, 356
 - risk management standards, 579
 - SR 11–7 (U.S. Board of Governors publication), 478
- Federal Reserve Board
- lender of last resort, 461
 - rules, relaxation, 38
- FICO score, 149–150
- Fidelity, asset management, 76
- Fiduciary account, usage, 38
- Finance, models, 484
- Financial Accounting Standards Board (FASB), 40
- FAS 157, 492
- Financial component (FC), 439
- Financial innovation, 619
- Financial innovators, regulation, 634–635
- Financial institutions
- bankruptcy, 17
 - capital requirement, 549f
 - disruption, 638
 - portfolio, 262
 - responses, 636–639
 - risk management, 18–19
 - risk–return alternatives, 516
- Financial instruments, 99
- creation, 452
 - valuing, models (usage), 491
- Financial performance, problem, 457
- Financial Services Authority (FSA), 459
- short selling ban, 102
- Financial Services Modernization Act (1999), 38
- Financial Stability Board, 519, 520, 575
- TCFD, 500, 506
- Financial Stability Oversight Council (FSOC), 578, 579
- Financial variables, daily percentage changes, 168–169
- Fink, Larry, 507
- FinTechs, services, 624
- Firm commitment, 32
- First order approximation, 678, 679
- First-order autocorrelation, 244, 245
- Fixed assets, 28
- Fixed coupons, usage, 377
- Fixed dollar payoff, providing, 119
- Fixed income arbitrage, 92
- Fixed-income arbitrage, 92
- Fixed-rate deposits, transformation, 111
- Fixed-rate loans, bank usage, 111
- Flash crash (2010), 635
- Flight to quality, 155, 352
- Flip-in, 36
- Floating forward rates, 665–666
- Floating-rate deposits
- bank usage, 111
 - transformation, 111
- Floor, put option series, 115
- Foreign currency options, profitability, 169
- Foreign currency risk (hedging), forward contracts (usage), 104
- Foreign exchange
- exposure, 120
 - rates, change, 558

- Foreign risk-free rate, 663–664
- Forward contracts, 103–104
 - calculations, 225–227
 - entry, 322
 - futures contracts, comparison, 107t
 - no-default value, 407
 - payoffs, 105f
 - position, closing out, 106
 - risk-neutral valuation, application, 220–221
 - software error, 107
 - usage, 104
 - valuation, 225, 663
- Forward exchange rates, assumption, 667t
- Forward price, volatility, 406
- Forward quotes, 103t
- Forward rates, 659
 - realization, 666t
- Forward trader, gains/losses, 107
- Four-index example, 263, 284–286
- French CAC 40, 122
- Frequency distribution, sampling, 435
- Front-end load, 78, 706
- Front office, separation, 645
- Front running, 83
- FTSE 100, 259, 267, 268
- Fundamental Review of the Trading Book (FRTB), 585, 591–596, 602
 - background, 586–587
- Funding decisions, 457
- Funding value adjustment (FVA), 394
- Fund managers, investments, 75
- Funds, 75–79, 84–88
 - assets, volatility (increase), 86
- Funds of funds, setup, 84–85
- Future cash flows, present value, 217
- Futures contracts, 104–108
 - consideration, 664
 - delivery, anticipation (absence), 105–106
 - delivery times/locations, 105
 - forward contracts, comparison, 107t
 - live cattle futures contracts, trading, 105–106
 - valuation, 663
 - VIX purchase, 167
- Future volatility (forecast), GARCH(1,1) usage, 185–188
- G-30 policy recommendations, 539–540
- Gamma, 327–329
 - beta, relationship, 699f
 - calculation, 400, 675
 - level, 334f
 - measure, 311
 - monitoring, 336
 - neutrality, 334
 - option (gamma), underlying stock price (relationship), 328f
 - zero level, 328
- GAP management, 314
- Gaussian copula, 202–203, 206, 208, 417
 - alternatives, 212
 - model, 204–205
 - cumulative joint probability distribution, 204t
- Gaussian quadrature, usage, 694
- General Data Protection Regulation (GDPR), 636
- Generalization, 282–284, 310–312
- Generalized autoregressive conditional heteroskedasticity (GARCH), 163
- Generalized autoregressive conditional heteroskedasticity, GARCH(1,1) model, 177–179
 - autocorrelations, 184t
 - daily volatility, 183f
 - estimation, 180–184
 - daily data, usage, 178
 - parameters, estimation, 181t
 - quality, determination, 184–185
 - selection, 179
 - usage, 185–188, 197
 - weights, 178–179
- Generalized Pareto (cumulative) distribution, 271
- General Motors bonds, sale (example), 100
- Generator matrix, 688
- Geometric mean, 82
- Germany, transition costs, 501
- Gerstner, Lou, 639
- Glasgow Financial Alliance for Net Zero (GFANZ), 500, 607
- Glass-Steagall Act (1933), 38–39
- Global Financial Crisis (GFC), 17, 19, 31, 129, 136
 - house value, decrease, 503
 - liquidity, tightness, 451
 - securitization, relationship, 145
- Global Investor Statement to Governments on the Climate Crisis (2021), 506–507
- Global macro hedge fund managers, 92–93
- Global macro, usage, 92–93
- Global Reporting Initiative, 509
- Global systemically important banks (G-SIBs), 356, 574–575, 575t, 611
- Going-concern capital, 568

- Gold
 - forward contract, bank purchase, 406
 - mining companies, hedging, 322
 - portfolio, summary, 320t
 - prices, 400
 - changes, 322
 - purchase, example, 100
- Goldman Sachs, 522–523
- Gone–concern capital, 568
- Google, IPO, 34, 35
- Google Wallet, services, 624
- Government National Mortgage Association (GNMA) (Ginnie Mae), 41
- Government-sponsored deposit insurance plan, impact, 61
- Great Britain Pounds (GBP)
 - purchase, long position, 105f
 - sale, short position, 105f
 - spot/forward quotes, 103t
- Greek letters
 - calculation, 319, 329, 332–333, 670t
 - RMFI software, usage, 333t
 - defining, 335
- Greeks, calculation, 324
- Green bond, 506
- Greenhouse gas emissions (economic sector), 499t
- Green light status, 358
- Greenspan, Alan, 156, 470
- Greenwashing, 509–510
- Group health insurance plans, 61
- Group life insurance, 51
- Guaranteed Annuity Option (GAO), 53
- Guarantees, 555
- Haircuts
 - ECB usage, 461
 - increase, 463
- Hammersmith and Fulham
 - interest rate swaps, 440–441
 - losses, 642
- Hashing, 623
- Hazard rates, 370–371, 383
 - average, 383t
 - bootstrapping, 380–381
 - default probability, calculation, 371
 - differences, reasons, 384–385
- Health insurance, 60–61
 - individual purchase, 61
- Heating degree days (HDD), 116–117
- Heavy-tailed distribution, normal distribution (comparison), 169, 170f
- Hedge and forget property, 322
- Hedge funds, 84–88
 - clawback clause, 85
 - dedicated shorts, 89–90
 - distressed securities, 90
 - fees, 85–86
 - global macro, usage, 92–93
 - high-risk investment, return, 87t
 - high-water mark clause, 85
 - hurdle rate, 85
 - implications, 84
 - leveraging–deleveraging cycle, impact, 470
 - long/short equity strategies, 89
 - managers, incentives, 86–88
 - money, losses, 94
 - performance, 93–94, 94t
 - profitability, reason, 707
 - strategies, 89–93
- Hedgers, speculator role (avoidance), 650–651
- Hedging, 489–490
 - error, 328f
 - facilitation, delta (calculation), 314–315
 - issues, 461–462
 - options, usage, 120–121
 - realities, 335
 - scheme, 326
- High-frequency traders (HFTs), problems, 635
- High-risk investment, return, 87t
- High-water mark clause, 85
- Histogram, usage, 260, 261f
- Historical data, usage, 8–9
- Historical default probabilities, 359–371
- Historical simulation, 257
 - approach, 288–289, 296
 - bootstrap method, 269–270
 - calculations, 258–262
 - total return stock indices, U.S. dollar equivalent, 259t
 - computational issues, 270
 - extensions, 265–270
 - losses
 - histogram, 261f
 - ranking, 261t, 266t
 - methodology, 257–263
 - model-building, contrast, 288–289
 - scenarios, generation, 260t
 - usage, 262
- Historical variables, shock, 349

- Homogeneity, 239
- Homogeneous expectations, 11
- Homogenous expectations, 11
- Horizon date, alternative scenarios (development), 413–414
- Housing bubble, creation, 149
- HSBC, reputational risk, 505
- Hunter, Brian, 87
- Hurdle rate, 85
- Hurricane Ida, impact, 499
- Hybrid funds, 77
- IBM shares, 244
 - purchase, example, 99
 - trading, 450
- ICE Clear Credit, 133
- Iksil, Bruno, 481–482
- Illiquid bonds, yields, 64
- Implied volatilities, 165–167, 487–490, 671
 - increase/decrease, 348
- In-arrears swap, 485
- Incentives
 - fees, 85
 - problems, 157–158
- Inception profits, recognition, 646
- Income statement, summary, 28t
- Incremental default risk charge (IDRC), 565–566
- Incremental ES, 248
- Incremental measures, 247–248
- Incremental risk charge, 564, 565–566
- Incremental value at risk, 247
- Independence projects, loss probability, 239
- Independent identical distributions, standard deviation, 244
- Independent normal distributions, sum, 244
- Index funds, 77
- Initial coin offerings (ICOs), 631
- Initial margin, 108, 132, 134
 - determination, 137
- Initial public offering (IPO), 33–35
 - Google IPO, 34, 35
 - shares, company sales, 33
- Instantaneous volatility increase (impact), GARCH(1,1) prediction, 188t
- Insurance, 443–445
 - policy, deductible (presence), 444
- Insurance companies
 - climate changes, impact, 501–502
 - expected cash flows, 50f
 - GAO offerings, 53
 - minimum capital requirements, 64
 - regulation, 65–66
 - risks, 64
 - role, 47
- InsurTech, 632–634
- Intel
 - call options, contract purchase (net profit), 114f
 - options, prices, 113t
 - shares, change, 113
- Intercontinental Exchange (ICE) trade, 117
- Interest leases and dividend component (ILDC), 439
- Interest-rate-dependent assets, long/short positions, 311–312
- Interest-rate-dependent securities, portfolio value, 310
- Interest rates
 - bank offering, 303t
 - changes, 297f
 - compounding frequencies, 655
 - correlations, 299t
 - example, 55–56
 - increase, 304t, 333
 - offsetting interest rate swaps, matched pair (expected exposure), 405f
 - options, 115
 - quotes, 657
 - risk, 293
 - swaps
 - currency swaps, contrast, 405
 - valuation, 666t
 - term structures, 293
 - exposure, 298
 - types, 294–296
 - zero rates, parallel shift, 310f
- Interest, risk-free rate, 221
- Intergovernmental Panel on Climate Change (IPCC), 498
- Internal credit ratings, 368
- Internal data, external data (contrast), 436–437
- Internal fraud, 434
- Internal historical data, inflation adjustment, 437
- Internal Loss Multiplier (ILM), calculation, 439–440
- Internal model-based approach, 543
- Internal models approach, 591–595
- Internal Ratings Based (IRB) approach (Basel II), 546, 549–551

- International Accounting Standards Board (IASB), 40
 - IAS 39, 492
- International Energy Agency (IEA), 498
- International Swaps and Derivatives Association (ISDA)
 - CDS trading contract, 373
 - master agreement, 135, 137, 396, 398, 540
- Internet banking, offering, 39
- Inverse chi-square distribution, 205
- Investment
 - alternative risk-return combinations, 5f
 - banking, 25, 31–36
 - beta, usage, 9–10
 - combination, 4
 - efficient frontier, 7f
 - expected return, 5f, 8
 - opportunities, 3–5
 - policies, disclosure, 84
 - portfolio, usage, 259t
 - return, standard deviation, 5f
 - risk-free investment, funds investment, 7f
 - riskiness, 4f
 - variables, independence, 11
- Investment grade bonds, 19
- Investors
 - advice, 38
 - climate risk concerns, 505–507
 - expected return, 10–11
 - protection, 579
 - risk aversion, 5
 - risk-free rate, usage, 11
 - risk/return, contrast, 2–5
- IPCC Sixth Assessment Report, shared
 - socioeconomic pathways, 498t
- Irrational exuberance (Greenspan), 156–158, 470–471
- Isserlis' theorem, 342–343
- Issue, bank underwriting (example), 32
- Ito's lemma, 387
- iTraxx Europe portfolio, 376
- Jett, Joseph, 485, 487, 641, 645
- Joint cumulative probability distribution, 204
- Joint distribution, copula model defining
 - method, 205f
- JPMorgan Chase (JPM), risk-weighted assets
 - reduction, 431, 481
- Jumpstart Our Business Startups (JOBS) Act (2012), 630
- Jump-to-default (JTD) risk, 591, 595
- Junk bonds, 19, 90
- Kahneman, Daniel, 524
- Kamakura (Moody's), 368, 388
- Kaufmann, Daniel, 508
- Kerviel, Jérôme, 120, 122–123, 432–433, 606
- Key risk indicators (KRIs), 441–442
- Kidder Peabody
 - losses, 641–643, 645
 - modeling mistakes, 484–485, 487
- Kirk, Stuart, 505
- KMV (Moody's), 368, 388
- Knock-out call option, 337
- Know-your-client (KYC) rules, 634
- Kray, Aart, 508
- Kurtosis, calculation, 341
- Late trading, 83
 - illegality, 83, 707
- LCH Clearnet, 141
- Leeson, Nick, 123, 641, 645
- Left tail (probability distribution), 273
- Lehman Brothers, 305, 571
 - bankruptcy, 16, 140, 649
 - default, 76, 534
 - derivatives portfolio, termination, 581
- Lenders
 - interest rate quotes, 657
 - online platforms, providing, 629
- Lenders of last resort, 461
- Lending, 629–631
- Less developed countries (LDCs), bank loans, 41
- Leverage
 - ratio, 568, 570–571
 - usage, limitation, 84
- Leveraging, 469–470, 470f
- Leveraging-deleveraging cycle, impact, 470
- Liabilities (maturities), interest rates
 - (increase), 304t
- Liar loans, 147
- Life assurance, 48
- Life expectancy
 - consideration, 55
 - variation, 56
- Life insurance, 47–51
 - beneficiaries, payout, 50
- Life insurance companies
 - abbreviated balance sheet, 63t
 - capital requirements, 63–64

- Likelihood function, 272
 - maximization, 184
- Likelihood measure, 182
- Linear products, 320–322, 321f, 324
- Lipper hedge funds database (TASS), 93
- Liquid assets, bank cushion, 465
- Liquidation cost, calculation, 453
- Liquidity, 305
 - backstop arrangements, 460–461
 - black holes, 466–472
 - funding risk, 457–466
 - horizons, 586–588
 - risk factors, allocation, 587t
 - importance, 140, 450
 - liquidity-adjusted VaR, 454–455
 - preference theory, 304
 - regulation, 471–472
 - risk, 449, 568, 571–573, 647–650
 - bank regulator management, 464
 - exposures/funding needs, bank monitoring/control, 464
 - shortage, 305
 - sources, 458–462
 - trading risk, 450–457
 - position, unwinding, 455–456
 - reduction, 453
- Liquidity-adjusted ES, calculation, 592–593
- Liquidity-adjusted VaR measure, 455
- Liquidity coverage ratio (LCR), 463, 572
- Litterman, Robert, 632
- Live cattle futures contracts, trading, 105–106
- Living wills, 579
- Ljung–Box statistic, 185
- Loan
 - default probability, estimation, 209
 - losses, avoidance, 40–41
 - portfolios (application), Vasicek’s model (usage), 208–212
 - replacement, 38
 - total value at risk (total VaR), consideration, 240
- Loan-to-value ratio, 149
- Log-likelihood function, maximum value, 274
- Lognormal distribution, 435
- London Interbank Offered Rate (LIBOR), 295
 - credit risk, 109
 - replacements, 108–109
- London Whale (trader risk), 431, 481–482
- Lone Star Funds, ABS CDO tranche purchase, 452
- Long call
 - position, 339
 - value, probability distribution, 339f
- Long-dated options, volatilities, 330
- Longevity, 56–57
 - bond, 57
 - derivatives, 57
 - risk, 56, 69
- Long position, 105f, 114
 - assumption, 103
- Long-run average covariance, 197
- Long-run variance rate, 174
- Long/short equity strategies, 89
- Long-tail risk, 58
- Long-Term Capital Management (LTCM)
 - liquidity risk, 648
 - losses/failure, 87, 351–352, 468, 471, 606, 643
 - positions, unwinding, 452
 - returns, report, 93–94
- Long-term funds, types, 77
- Long-term interest rates, options, 467
- Long-term needs, short-term funding (impact), 649
- Long-term volatility, 182
- Lookback options, 119
- Loomis, Carol, 84
- Loss (losses)
 - calculation, 381t
 - correlations, 609t
 - examples, 642–643
 - frequency distribution, 434–435
 - funding, order, 134
 - histogram, 260, 261f
 - probability density function, 549f, 604f
 - ranking, 261t, 266t, 268t
 - realization, 338t
 - risk, 343
 - sampling, 418
 - severity distribution, 435
 - severity/frequency, 434–438
- Loss Component (LC), 440
- Loss distribution, 235, 258
 - 99th percentile, estimation, 264
 - calculation, loss frequency/severity (impact), 436f
 - characteristics, 605t
 - shapes, 604–605
- Loss/gains outcomes, calculation, 236
- Loss given default (LGD), 416–417, 552, 591
- Loss ratio, 59
 - property-casualty insurer calculation, 59

- Macaulay, Frederick, 308
- Macaulay's duration, 307
- Machine learning (ML)
 - methodologies, 621f
 - models, 621
 - usage, 620–622
- Macroeconomic conditions, impact, 208, 384–385
- Macroeconomic variables, projections, 356
- Madoff, Bernard (Ponzi scheme), 122
- Maintenance margin, 101
- Managed futures, 93
- Margin, 131–133, 468
 - calculation, 130–131
 - requirement, 132
- Marginal distribution, 202
- Marginal ES, 248
- Marginal measures, 247–248
- Margin value adjustment (MVA), 394
- Market
 - derivatives markets, 102–103
 - discipline, 545
 - Basel II pillar, 556–557
 - emerging markets, 92
 - funds, outperformance, 80
 - liquidity
 - measurement, 453–454
 - measures, 457
 - outguessing, 644
 - price, 40
 - return, expected return (relationship), 12, 12f
 - timing, 83
 - transparency, 649–650
 - variables, 18, 226, 262
 - extreme movements, 349
 - movements, adjustment (losses ranking), 268t
 - variable, volatility
 - estimation, 176
 - scaling, 266–268
- Market maker, swap quote creation, 112t
- Market portfolio
 - return, 9, 12
 - example, 10
 - risk-free rate, example, 10
- Market risk
 - economic capital, 602
 - losses/gains, one-day standard deviation, 602
 - loss probability density function, 604f
- Marking to market, 542, 691
- Marking to model, 492, 646
- Markowitz, Harry, 279
- Markowitz portfolio theory, 234
- Mark-to-market valuation, 454–455
- Mark-to-market values, 542
- Maturities, interest rates, 55–56
- Maturity adjustment (MA), 552–553
- Maturity date, 112
- Maximum likelihood methods, 179–185
- McFadden Act, 27
- Media and Technology Group (TMTG), SPAC
 - action, 37
- Mercury Asset Management, 605
- Merger
 - arbitrage, 90–91
 - avoidance, 36
- Merrill Lynch, losses, 156
- Metallgesellschaft (MG), hedging, 462–463
- Metro do Porto (Mdp), funding cost
 - reduction, 521
- Mezzanine ABS CDOs, 152
 - trading opportunity, 159
- Mezzanine tranche, 150–153
 - rating issue, 155
- Microsoft
 - hedging, 120–121
 - shares, portfolio, 281
- Middle office, separation, 645
- Mid-market price, fair price (equivalence), 453
- Minimum capital requirements, 545
- Model-building
 - approach, 279, 289
 - methodology, 280–282
 - historical simulation, contrast, 288–289
 - missteps, 493–494
- Model documentation, 479
- Model risk
 - management, 477
 - regulatory guidance, 478–484
 - reduction, 480
 - SR 11–7 definition, 478
- Models
 - blind trust, avoidance, 646
 - development, 479–480
 - parameters, 484
 - simplicity, 484–487
 - validation, 480, 482–483
- Model validation, 480
- Modified duration, 307–308
- Money, borrowing ability, 458, 460
- Money center banks, 26
- Monotonicity, 239

- Monte Carlo simulation
 - refinement, 419
 - trials, 398, 421
 - usage, 396, 418, 420, 423, 435, 602
 - weighted Monte Carlo simulation, 491–492
- Moral hazard, 31, 61–62, 444
 - examples, 61
- Mortality
 - rates, increase/decrease, 558
 - risk, 56–57
 - tables, 53–56
 - data (2019), 54t
- Mortgage
 - lenders, good faith determination, 579
 - lending, 147
 - originators, incentive, 157
- Mt. Gox (bitcoin issue), 627–628
- Multibank holding company, 27
- Multivariate copulas, 207
- Multivariate normal distributions, 199–201, 207
 - usage, 200–201
- Mu (μ), selection, 275–276
- Mutual funds, 75–79
 - costs, 77–79
 - good performance, consistency, 81t
 - holding, average total cost per year, 78t
 - investment policies, disclosure, 707
 - open-end mutual funds, asset growth, 76t
 - regulation, 83
 - returns, problems, 82
 - valuation, 77
- Naked CDS positions, 373
- National Association of Insurance Commissioners (NAIC), 65
- National Westminster Bank, losses, 643, 646
- Natural disasters, 58
- Natural gas derivatives, 117–118
- Natural logarithm function, 218
- Natural probability distribution, 435
- Negative binomial distribution, 418
- Negative feedback traders, actions, 466–468
- Negative skewness, 11
- Net asset value (NAV), 77
 - calculation, 79, 83
 - daily calculation, 84
- Net interest income
 - constancy, 28–29
 - management, 303–305
- Net interest margin, fluctuations (occurrence), 303
- Net replacement ratio (NRR), 541–542
- Net stable funding ratio (NSFR), 463
 - ASF factors, 572t
- Netting, 130, 135, 139f, 540–542, 548
- Network for Greening the Financial System (NGFS), 498
- New York Stock Exchange (NYSE), trading
 - focus, 130
- Nikkei 225, 259, 267, 268
 - futures quotes, arbitrage, 123
- Noncontributory policy, 51
- Non-dividend-paying stock
 - American put, binomial tree, 674f
 - long forward contract, 220–221
- Nonfinancial corporations, lessons, 650–651
- Noninvestment grade bonds, 19, 90
- Nonlife insurance, 47
- Non-linear derivatives, 279
- Non-linearity, 328f
 - handling, 288
- Nonlinear portfolio delta neutral protection, 323
- Nonlinear products, 323–327
- Non-modelable risk factors, 594
- Non-normal distributions, 287
- Nonparallel yield curve shifts, 312–315
- Non-stochastic term, 335
- Nonsystematic risk, 9, 14–15, 385
- Non-traditional derivatives, 115–119
- Non-U.S. legislation, 580–581
- Normal distribution
 - assumption, 607
 - heavy-tailed distribution, comparison, 169, 170f
- NORMSDIST/NORMSINV, usage, 209, 219, 243
- Northern Rock, 449, 461, 571
 - assets, financing inability, 305
 - failures, 17, 649
 - shares, floating, 458–459
- Notice of intention to deliver, issuance, 106
- Notional principal, 372, 379
- Obama, Barack, 60, 564
- Objective probabilities, subjective probabilities
 - (contrast), 360–361
- Observations
 - likelihood, 180
 - weighting, 265–266
- Off-balance sheet items, 536, 570
- Offer price, 130
- Office of Financial Research (OFR), 578, 579

- Office of Thrift Supervision (OTS),
elimination, 578
- Oil derivatives, 117
- Olayan Group LLC, 576–577
- On-balance sheet assets
principal, 538
risk weight, 536
- On-balance sheet exposures, 570
- On-balance sheet items, risk weights, 537t
- One-basis point change, impact, 302
- One-day VaR
back-testing, 250
calculation, 247
- One-factor copulas, 212
- One-factor Gaussian copula model, 693–694
- One-factor model, 201
- One-year loss distribution, economic capital
(calculation), 600f
- One-year risk-free interest rate, impact, 13
- On-the-run bonds, purchase, 649
- Open banking, 628–629
- Open-end fund, 76
- Open-end mutual funds, asset growth (United States), 76t
- Open outcry system, 130
- Operating ratio
property-casualty insurance company
calculation, 60t
property-casualty insurers calculation, 59
- Operational risk, 29, 431
assessment, result, 443
Basel II, impact, 555–556
capital allocation, 442–443
correlation, 608
defining, 433–434
economic capital, 603
e-mails/phone calls, impact, 442
losses, prevention, 440–442
loss probability density function, 605f
types, 434
- Options, 112–115
delta-neutral portfolio, gamma, 335
gamma, underlying stock price
(relationship), 328f
interest rate options, 115
in-the-money, 169
markets, trades (types), 114
out-of-the-money option, 114
premium, 112
prices (Intel), 113t
synthetic creation, 467–468
types, 119
usage, 120–121
value, 114
obtaining, expected payoff (discounting), 219
vega neutral option, 330
writing, 114
- Orange County, losses, 641, 643, 650
- Organized trading facilities (OTFs), 136, 580
- Originate-to-distribute model, 41–42, 460
- Outcomes analysis, 482
- Out-of-state acquisitions, prohibition, 27
- Out-of-the-money option, 112, 114
writing, 94
- Outside-model hedging, 489
- Outstanding loans, VaR percentage, 612–613
- Over-fitting, 493
- Overnight rate, 294
- Over-parameterization, 494
- Over-the-counter (OTC) contracts, 102–103, 117
- Over-the-counter (OTC) derivatives, 375,
395, 536
changes, 136–141
clearing, CCP approval, 136
electronic platform trading, requirement, 136
initial margin, attraction, 139
transactions, 133, 697
- Over-the-counter (OTC) derivatives market, 129
clearing, 131–136
exchange-traded markets, convergence, 141
liquidity, importance, 140
margin, 131–133
post-crisis regulatory changes, 136–138
rehypothecation, 140–141
- Over-the-counter (OTC) markets, CCP role, 133f
- Over-the-counter (OTC) trades, central trade
repository report (requirement), 137
- Over-the-counter (OTC) weather derivatives, 116
- Page, Larry, 35
- Partial derivative
calculation, 611
usage, 320
- Partial duration, 312–314, 314t
- Partial expected shortfall, 593
- Partnership for Carbon Accounting Financials
(PCAF) forum, 502
- Passive investors, 90
- Patient Protection and Affordable Care Act, 60–61
- Paulson, John, 523

- Payment Services Directive 2 (PSD 2), 629
- Payment systems, 624–628
- Payoff, basis, 119
- Payout
 - occurrence, 55–56
 - present value, 63
 - total present value, 56
- PayPal, services, 624
- Peak exposure, 398
- Peer-to-peer (P2P) lending, 629–630
- Penalty-free withdrawals, 52
- Pension plans, 43–44, 47, 66–69
 - longevity risk, concern, 69
 - perfect storm, creation, 68
- Percentage change, 282
- Percentiles
 - analytical results, 339–343
 - weight functions, 242f
- Pérez de Cuéller, Javier, 508
- Peripheral variables, core variables (contrast), 350
- Photography, digitization, 637
- Physical assets, damage, 434
- Physical risks, contrast, 500
- Physics, models, 484
- Pipeline risk, 355
- Plain vanilla derivatives, 103–115
- Plain vanilla interest rate swap, 110–111, 110f
- Plain vanilla options, linear combinations, 591
- Plain vanilla swap, 485
- Poison pills, examples, 36
- Poisson distribution, 418
- Policy limit, 444
- Ponzi scheme, 122
- Porter, Michael, 526
- Portfolio
 - asset price dependence, 333
 - beta, 514
 - examination, 12
 - portfolio manager usage, 13
 - change, mean/standard deviation estimates, 698
 - convexity, 311
 - dependence, 330
 - expected return, market return (relationship), 12, 12f
 - expected shortfall (ES), calculation, 241
 - financial institution instruments, 225–226
 - gain, 236
 - gamma neutral, 328–329
 - hedging, 314
 - immunization, 311–312
 - interest rate term structures, exposures, 298
 - losses, variance, 285
 - market portfolio, return, 10
 - mean/standard deviation, dependence, 515f
 - one-day 99% VaR, 280
 - partial durations, 314t
 - time decay, 331
 - value at risk (VaR), calculation, 240, 241
 - vega, 329–330
 - volatility scaling, 268–269
- Portfolio loss, 313
 - normal distribution, 243
 - standard deviation, 268
 - VaR calculation, 243
- Portfolio value, 319
 - change, standard deviation, 284
 - daily changes, 245
 - gain, probability distribution, 237f
- Position, unwinding, 452, 455–456
- Positive alpha, 81
- Positive equity, 17
 - capital, 568
- Positive feedback
 - traders, 466–469
 - trading, existence (reasons), 467–468
- Positive gamma portfolio, 340
- Positive-semidefinite, 198–199, 201
- Positive skewness, 11
- Post-crisis changes, 563
- Power law, 170–172
 - equivalence, 272–273
 - usage, 443
- Predatory trading, 468
- Premium payments, consideration, 56
- Pricing model, success (factors), 492–493
- Prime brokers, rule, 88
- Prince, Chuck, 471
- Principal, cash flow example, 110t
- Principal components analysis (PCA), 299–302, 685
- Probability density function, 211
- Probability distribution (PD), 193, 419t
 - assumptions, 341–342
 - calculation, 225
 - equation, 208
 - estimation, 211
 - impact, 235f
 - left tail, 273
 - loss probability density function, 549f
 - occurrence, impossibility, 237

- Probability distribution (PD) (*continued*)
 - shape, 420f
 - tail, estimation, 272
 - WCDR, relationship, 552t, 554t
- Process management, 434
- Procter and Gamble (P&G), 520, 645
 - 5/30 swap, 121–122
 - losses, 643
- Profit
 - easy profits, caution, 647
 - realization, 338t
- Profit & loss (P&L)
 - attribution, 594–595
 - decomposition, 490
- Property–casualty insurance, 43, 57–59
 - companies
 - abbreviated balance sheet, 64t
 - operating ratio calculation, 60t
- Property–casualty insurers
 - loss ratios, 65
 - ratio calculations, 59
- Property damage, natural disasters, 58
- Proportional bid–ask spread, mean/standard
 - deviation, 454
- Proprietary trading, 578
- Public offering, 32
- Put options, 112
- Put, purchase/sale, 114
- p*-value, 594

- Qatar Holding LLC, 576–577
- q*-percentile, 263
- Quantitative Impact Studies (QIS), 545
- Quantum Fund, 93

- Random samples generation, multivariate normal
 - distributions (usage), 200–201
- Ratings transition matrices, 414–415
 - probabilities, expression, 415t, 416t
- Ratios, property–casualty insurer calculations, 59
- Real-world default probabilities, risk-neutral
 - default probabilities (contrast), 388
- Real-world expected payoff, 222
- Rebalancing, 324–325
- Recovery account, 85
- Recovery rates, 371–374, 421
 - dependence, 372
 - expectation, 380
- Red light scenario, 357
- Red light status, 357–358

- Reference entity, 372–374
- Reference rates, 294–295
- Regime change, 494
- Regional banks, role, 26
- Regression error, 8
- RegTech, 634–636
- Regulation/compliance, impact, 634–636
- Regulatory arbitrage, 157
- Rehypotheication, 140–141
- Reinforcement learning (RL), 622
 - models, 621
- Reinsurance, 57, 62–63
 - contracts, forms, 63
- Reintermediation, 619
- Relative valuation, 491
- Relative value strategy, 92
- Renaissance Technologies Corp., 85
- Replacement ratio, 542
- Repo rates, 295–296
- Representativeness, 525
- Reputational risk, 505, 603
- Required stable funding (RSF), 464
 - factor, 572, 573t
- Reserve requirements, 463
- Residential mortgages, bank assets, 554
- Residual risk add-on, 591
- Retail banking, lending rate, 25
- Retail customers, bank loans (default probability
 - estimation), 209
- Retail deposits, funding source, 460
- Retail exposures, 553–554
- Retinal scans, usage, 625
- Return
 - arithmetic mean, 82
 - distribution, moments (examination), 10–11
 - example, 10
 - expected return, 5f
 - investor focus, 11
 - multiple, 9
 - risk, contrast, 2–5
 - standard deviation, 5f, 7
 - investor attention, 10–11
- Return on equity (ROE), 29
- Return on risk-adjusted capital (RORAC), 613
- Reverse stress testing, 352–354
- Rho (μ), 332, 333t, 671, 675
- Riegel-Neal Interstate Banking and Branching
 - Efficiency Act, 27–28
- Right-way risk, 401
- Ripple effect, creation, 534

- Risk
 - appetite, 32, 514–520
 - assumptions, constant level, 425
 - budgeting, 249
 - capital, 599
 - categorization, 601f
 - charge, change, 588
 - credit risk, 109
 - culture, 520–524
 - decomposition, 18
 - expected return, trade-off, 2–3
 - factors, 258, 282, 586, 591
 - allocation, 587t
 - identification, 524–526
 - insurance company risk, 64
 - interactions, 606
 - leveraging, 222
 - limits, 237, 642–645
 - margin period, 397
 - measures, calculation, 288, 296–299
 - mitigation, effectiveness, 355
 - quantification, 3, 519
 - relative importance, 605–606
 - return, contrast, 2–5, 14–17
 - types, loss distributions (characteristics), 605t
 - weights, 287–288, 536
 - adjustments, simple/comprehensive approaches, 548
- Risk-adjusted return on capital (RAROC), 599, 612–613
- Risk control and self-assessment (RCSA), 441–442
- Risk-free interest rate, 323
- Risk-free investment
 - denotation, 6
 - funds investment, 7f
- Risk-free rate (RFR), 12, 109, 378
 - example, 10
- Risk-free zero-coupon bond, yield, 404
- Risk management, 18–19
 - challenges, 120–123
 - company policies, 539
 - mistakes, avoidance, 641
- RiskMetrics, 234
- Risk-neutral default probabilities, real-world
 - default probabilities (contrast), 388
- Risk-neutral valuation, 219–224
 - application, 220–222
 - calculations, 225–227
 - principles, 323
 - result, 219
 - usage, 224–225, 386
- Risk-neutral world, defining, 219
- Risk-return trade-offs, 1, 75
 - improvement, 5
- Risk takers, classification, 581
- Risk-weighted amount, 536
- Risk-weighted assets (RWAs), 544
- Risk-weighting adjustments, absence, 570
- Risky investments
 - efficient frontier, obtaining, 6f
 - portfolio, 8
- RMFI software, usage, 257, 333t, 695, 771
- Robertson, Julian, 84, 92
- Robo-advisers, impact, 631
- Robotics process automation (RPA), 620
- Rogue trader
 - insurance, 444–445
 - losses, 122
- Rolling over, 305
- Royal Dutch Shell, Millieudedefensie lawsuit, 504
- Rusnak, Jhon, 641
- Russia, debt (default), 352, 451
- Sandbox programs, 634
- Santander, rail deal (lawsuit), 521–522
- Sarbanes-Oxley Act (2002), 445–446
- Sasson, Steve, 637
- Savings institutions, US ranking, 26t
- Scenario analysis, 217, 224, 338, 347, 438
 - usage, 224–225, 645
 - valuation, contrast, 385–386
- Scenarios
 - completion, 350–352
 - generation, 347–354, 438
 - management generation, 349–350
 - probabilities, assessment, 360
 - regulator selection, 356–357
- Schloss, Walter J., 84
- Second order approximation, 678, 679
- Second partial derivative, usage, 327
- Sector neutrality, 89
- Secured overnight funding rate (SOFR), 109–111, 294
- Securities financing transaction exposures, 570
- Securities trading, 37
- Securitization, 149–155, 460–461, 595
 - global financial crisis, relationship, 145
- Seller's market, 451
- Semiannual compounding, usage, 307–308

- Senior management
 - questions, brainstorming, 350
 - strategy development, 464
- Senior tranche, 150
- Sensitivity, calculation, 307–308
- Services component (SC), 439
- Shadow banks, 634
- Shared Socioeconomic Pathways (SSPs), 498
- Share-for-share exchanges, 90
- Shares
 - company sales, 33
 - purchase, 99, 101t
 - redemption, 84
- Shocks, 586
- Short call, value (probability distribution), 340f
- Short position, 103, 105f, 106
- Short sales, 100–102
 - cash flows, 101t
- Short selling ban, 102
- Short-squeezed investor, 100
- Short-term debt instruments, sales, 458–459
- Short-term funding, usage excess (avoidance), 649
- Simons, James, 85, 86
- Simple approach (collateral risk weights adjustment), 548
- Simulated losses (monitoring), volatility results, 269t
- Single derivatives transaction, positive value, 403–404
- Single forward transaction, 405–407
- Skewness, 11, 342
 - positive skewness, 11
- Social sustainability, 509
- Société Générale (SocGen), 120, 431, 606, 641
 - derivatives risk, 122–123
 - losses, 643
- Software, errors, 107
- Solar Radiation Management, climate change response, 500
- Solvency, 449
- Solvency capital requirement (SCR), 557–558
- Solvency I, 66
- Solvency II, 66, 533, 557–558
- Solver (Excel), usage, 274
- Soros, George, 84, 92, 93
- Sovereign debt
 - climate change, impact, 504
 - naked CDS positions, 373
- Sovereign exposures, 551–553
- Sovereign wealth funds, 43
- Special Purpose Acquisition Companies (SPACs), 34
 - bubble, 34
 - Trump, relationship, 37
- Special purpose vehicle (SPV), 150
- Specific risk charge (SRC), 543
- Spectral risk measures, 242
- Speculative grade bonds, 19
- Spider (SPY), 79
- Spinning, 33
- Spot exchange rate, applicability, 104
- Spot market, GBP purchase, 103
- Spot quotes, 103t
- SR 11-7 (U.S. Board of Governors publication), 478
- Stable funding, amount (calculation), 573
- Standard deviation (σ), 5f, 243, 298, 299t
 - change, probability, 172
 - estimate, 172
 - factor scores, standard deviation, 300t
 - investor attention, 10–11
 - move, 164
 - one-day standard deviation, 602
- Standard Initial Margin Model (SIMM), 137–138, 343, 697
 - ISDA proposal, 698
- Standardized approach, 431
- Standardized approach (Basel II), 546–548, 554, 588–591
 - usage, 577–578
- Standardized approach-counterparty credit risk (SA-CCR), usage, 577–578
- Standardized capital charge, 567t
- Standardized measurement approach (SMA), 439–440, 577
- Standard & Poor's 500 (S&P500), 182f, 268
 - futures contract, 664
- GARCH(1,1)
 - daily volatility, 183f
 - model, parameters (estimation), 181t
 - total annual return, 516t
 - volatility term structure, GARCH(1,1) prediction, 188t
- Static options, replication, 336, 337f
- Statistics, methodologies, 621f
- Sterling, futures contracts, 107
- Sterling interest rates, movements, 440–441
- Stock
 - beta, 220

- indices, actively managed funds
 - (outperformance), 80
- investment, one-year return, 3t
- selection, 13
- Stock market crash (1987), 471, 526
- Stock prices, 323
 - changes, absence, 333
 - decrease, 325
 - expected growth rate, 221
 - probability distribution,
 - calculation, 225
 - real-world probability, 221–222
 - sequences, 172, 173t
 - volatility, 323
- Stop-loss rules, 467
- Strategic risk management, 526–527
- Stressed daily volatility, 588
- Stressed ES, 263, 286
- Stressed market conditions, 195
- Stressed measures, 286–287
- Stressed VaR, 263, 286, 564–565
 - calculation, 564
- Stressed VaR/stressed ES, 263
 - calculation, 348
- Stress scenarios, consideration, 359–360
- Stress testing, 347–348, 465, 645
 - committee, input, 359
 - importance, 350
 - programs, usage, 354
 - recommendations, 354–356
 - regulation, 354–357
 - results, 358–361
 - rigorousness/comprehensiveness, 354
 - VaR calculations, integration, 358–360
- Strike price, 112–113, 114f, 118, 323, 488
- Structured products, 119–120
- Student's t -copula, 205–206
- Student's t -distribution, 206
- Stumpf, John, 518
- Subadditivity, 239
 - condition, 242
- Subjective default probabilities, 384
- Subjective probabilities, objective probabilities
 - (contrast), 360–361
- Subprime crisis (2007), 458
- Subprime first mortgage, 146
- Subprime mortgages, 40, 146
 - losses, 643
- Subprime securitizations, ABS/ABS CDO/CDO
 - creation, 154f
- Sunk cost bias, 526
- Supervised learning models, 621
- Supervisory review, 545
 - Basel II pillar, 556
- Surplus premiums, 49
- Survival probabilities, 690t
- Sustainability, 497, 508–509
 - pillars, 509
- Sustainability Accounting Standards Board (SASB),
 - founding, 509
- SwapClear, 133, 141
- Swap execution facilities (SEFs), 136, 578
- Swaps, 109–111
 - agreement, 111
 - duration, 110t
 - entry, 408f
 - options, 115
 - plain vanilla interest rate swap, 110f
 - quotes, market maker creation, 112t
 - valuation, 665
- Swaptions, 115
- Swiss average rate overnight (SARON), 109,
 - 294–295
- Swiss Re, 62
- Synthetic CDOs valuation, 693
- Synthetic put options, creation, 648
- Systematic risk, 3, 13, 14, 384–385
 - investor trade-off, 9
- System failures, 434
- Systemically important financial institutions (SIFIs), 574–575, 579
 - failure, 581
- Systemic risk, 534–535
- Tail dependence, 206, 212
- Takeover, avoidance, 36
- Task Force on Climate-Related Financial Disclosures (TCFD), 500, 504, 506
- Taylor series expansions, 333–335, 339, 677
 - results, 679
- T -Day VaR to One-Day VaR (ratio), 245t
- t -distribution, 604
- Teaser-interest-rate period, 148
- Technological advances, impact, 620–624
- Telephone banking, offering, 39
- Temporary life insurance, 48
- Tencent, services, 624
- Term life insurance, 48
 - break-even premium, 55–56
- Term structures, 296–297, 589

- Theta, 331–332
 - calculation, 675
 - variation, 331f
 - patterns, 332f
- Third order approximation, 678
- Tier 1 equity capital (CET1), 568–570
- Tier 2 capital, 568
- Tier I capital, 539
- Tier II capital, 539
- Time decay, 331
- Time horizon, 238, 244–246, 283
 - usage, 42
- Tokyo average overnight rate (TONAR), 109, 294
- Top-down approach, 601
- Total capital, equation, 544, 546
- Total credit risk economic capital, 609
- Total economic capital, allocation, 610
- Total enterprise-wide economic capital, 609
- Total expected loss, 396
- Total expense ratio, 78
- Total loss-absorbing capacity (TLAC), 575
- Total loss, experience, 435
- Total operational risk economic capital, 609
- Total return index, 260
- Total return stock indices, U.S. dollar
 - equivalent, 259t
- Total Tier 1 capital, 568
- Tourre, Fabrice, 442, 523
- Traders, actions, 647
- Trades
 - bid-ask spread cost, relationship, 455
 - size, relationship, 450, 451f
 - understanding, 650
- Trading
 - patterns, monitoring, 487–488
 - room, management, 645–647
 - strategy, examination, 648–649
- Trading book, 40, 157, 542
 - banking book, contrast, 595–596
 - mark-to-market valuation, 454–455
 - positions, liquidation ability, 458, 460
 - review, 585
- Traffic light options, 357–358
- Tranche
 - losses, 693
 - loss payments, 695
 - principal, 694
 - risks, 154–155
 - spread, 693
- Transactions, mark-to-market values, 542
- Transition matrix, usage, 414–415, 421–422
- Transition risks, contrast, 500
- Translation invariance, 239
- Transparency, importance, 452
- Treasury bills/bonds, investor earning, 294
- Treasury Department, profit center role
 - (caution), 651
- Treasury rates, 294, 661
 - decline, 68
 - factors, 301f
- Treasury securities, holdings, 458
- Treasury yield (TSY) price, calculation, 121–122
- Treasury zero rates, determination, 661
- Trend trading, 467
- Triangular distributions, 202f
 - standard normal distribution, presence, 203t
- Troubled Asset Relief Program (TARP), 581
- True probability, 251
- Trump, Donald (SPAC relationship), 37
- Tversky, Amos, 524
- Two-asset case, 281
- Two-tailed test, 164, 251
- Two-way zero threshold collateral agreement, 397
- Type I errors, 369
- Type II errors, 368
- UBS, losses, 156, 643
- Uncleared trades, 137
- Uncollateralized over-the-counter option, sale, 404
- Unconditional default probabilities, 370, 690t
- Unconditional (cycle-neutral) model, 603
- Unconditional probability, 272
 - density function, evaluation, 275
- Underlying asset, price, 331f
- Underlying stock price, option gamma
 - (relationship), 328f
- United States
 - banking
 - structure, 26
 - system, consolidation, 28
 - Environmental Protection Agency (EPA),
 - greenhouse gas estimates, 498–499
 - insurance company regulation, 65–66
 - open-end mutual funds, asset growth, 76t
 - real estate prices, S&P/Case-Shiller
 - Composite-10 Index, 146f
 - Treasury rates, factor loadings, 300t
- United States Dollars (USD)
 - GBP exchange rate, spot/forward quotes, 103t
 - net revenue, Microsoft conversion, 120–121

- United States housing market, 146–149
 - bubble, collapse, 148–149
 - crisis, lessons, 158
 - lending standards, relaxation, 146–148
- Univariate standard normal
 - distribution, 200
- Universal life (UL) insurance, 50
- Unsupervised learning models, 621
- Uptick rule, 101–102
- Valuation, 217
 - scenario analysis, contrast, 385–386
- Value adjustments (xVAs), 393, 407–408
- Value at risk (VaR), 233
 - accuracy, 263–264
 - aggregating, 249
 - back-testing, 594
 - difficulty, 252
 - calculation, 235, 243, 273, 302, 418, 697
 - confidence level, impact, 237
 - investment portfolio, usage, 259t
 - probability distribution, impact, 235f
 - stress testing, integration, 358–360
 - confidence level, 250, 264
 - definition, 234–235
 - historical perspectives, 233–234
 - hybrid approach, 608–609
 - measure, 543
 - methodology, 280–282
 - model
 - back-testing, 251–252
 - development, 482
 - one-day VaR, calculation, 247
 - parameters, selection, 242–247
 - problems, 237–238
 - procedure, extensions, 286–287
 - stressed VaR, 263
- Vanna, 335
- Variable life (VL) insurance, 50
- Variables
 - functions, 678–679
 - involvement, scenarios, 348–349
 - probability distribution, 273
 - stressing, 348
- Variable-universal life (VUL) insurance, 50–51
- Variance-covariance approach, 279
- Variance-covariance matrix, 198t, 284
- Variance rate, 165, 172
 - expected path, 186f
- Variation margin, 108, 133, 137
- Vasicek's model, usage, 208–212, 384–385, 416–417
- Vasicek's result, proof, 210–211
- Vega, 329–331
 - calculation, 400
 - monitoring, 336
 - neutrality, 330
 - risk, 589
 - variation, 331f
 - vega neutral portfolio, 330
- Vendor models, 483–484
- Vickers, John, 580–581
- VIX Index, 165, 167, 167f
- Volatility
 - adjustment, 269t
 - asset prices, relationship, 218–219
 - causes, 166
 - changes
 - absence, 224
 - impact, 176, 188
 - computation, data, 173t
 - definition, 163–167
 - estimation, 176
 - EWMA, usage, 267t
 - incorporation, variable-by-variable basis, 268
 - increase, 333
 - monitoring, 269t
 - scaling, 268–269, 348
 - surface, 487, 488t
 - term structures, 187–188
 - trader implication, 647
 - weighting schemes, 174–175
- Volatility-scaling procedures, 275
- Volcker, Paul, 578
- Volcker rule, 99, 578
- Vomma, 335
- Voting structure, change, 36
- Warehousing risk, 355
- Waterfall, 151–152, 151f
- Watson Regulatory Compliance (IBM), 636
- Wealth management, 631–632
- Wealth transfers, 69
- Weather derivatives, 89, 116–117
- Weatherstone, Dennis, 233–234
- Weighted expected shortfall (WES), 593–594
- Weighted Monte Carlo simulation, 491–492
- Weighted sensitivities, 287–288
- Weighting scheme, 196
- Wells Fargo, cross-selling, 517–519

- West Texas Intermediate (WTI) crude oil, 117
- Whole life insurance, 48–49
 - expected cash flows, 50f
 - surrender, 49
- Wholesale banking, lending rate, 25
- Wholesale deposits, funding source, 460
- Within-model hedging, 489
- With-profits endowment life insurance policy, 51
- Working capital, calculation, 369
- Worldwide Governance Indicators (WGI), 508
- Worst case default rate (WCDR), 209, 212, 416–417, 550–554
 - probability distribution, relationship, 552t, 554t
- Worst case outcome, interest, 225
- Writing an option, 114
- Wrong-way risk, 401–402

- Xerox, downgrade triggers (impact), 398
- Xi (ξ)
 - estimation, 272
 - values, 274

- Yellow light scenario, 357–358
- Yellow light status, 358
- Yield
 - changes, 308
 - providing, 670t

- Yield curve, 293
 - bucketing approach, usage, 315f
 - parallel shift, 312, 348
 - rotation, 314f
 - slope, twist/change, 300
- Young, Peter, 606

- Zero-coupon bond
 - company ownership, 424
 - outcomes, 424t
 - issuance, 404
 - usage, 386
 - yield curve, 92
- Zero-coupon interest rates, 293, 296
- Zero-coupon yield curve, 312f, 659
 - change, 310, 312, 313f
- Zero curve, 662f
 - trader calculation, 647
- Zero emissions, commitment, 506
- Zero-mean normally distributed variables, usage, 342–343
- Zero-mean unit-variance
 - distributions, 207
- Zero rates, 659, 660t
 - forward rates, 660t
 - parallel shift, 310f
- Z-score (Altman), 368–369

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.