2025

FRM®

EXAM PART I

Valuation and Risk Models

GARP®

FRM® | Financial Risk Manager

GARP®

FRM® | Financial Risk Manager

2025

FRM®

EXAM PART I

Valuation and Risk Models

Pearson

Pearson Education, Inc., 221 River Street, Hoboken, NJ 07030
A Pearson Education Company
www.pearsoned.com

Printed in the United States of America

00040348-00000002 / A103001322223

EEB/KC

Pearson

ISBN 10: 0-13-538972-0
ISBN 13: 978-0-13-538972-0

Contents

PREFACE

On behalf of GARP's Board of Trustees, the FRM advisory committee, and GARP's FRM professional certification program staff. I want to thank you for your interest in and support of the FRM program.

The program's first offering in 1997 saw just over 100 candidates sit for the exam. During the past 27 years, hundreds of thousands of professionals have studied for and taken the FRM exam, with it now being the world's leading financial certification program.

The dynamic nature of the FRM program's curriculum means that it regularly and quickly responds to changes in the global financial marketplace. This ensures that its content and reach always address the risks and challenges of a fast-changing, complex, and globally connected financial system.

For example, for 2025, after much discussion and consideration, the FRM advisory committee made material changes to the program's 2025 market risk measurement and management content. The result is that about half of the subject readings in Market Risk Measurement and Management were updated.

But maintaining a current and highly relevant curriculum is not the sole focus of GARP's professional staff. GARP has focused considerable time and resources during the past year developing tools to assist a candidate in his or her exam program preparation. In addition to providing current content, a primary objective of ours is to ensure as much as possible that a candidate is making the best use of his or her valuable time in preparing for the exam.

In this regard, GARP offers FRM Part I candidates an electronic platform called GARP Learning. GARP Learning is a streamlined digital learning program that can be accessed via a mobile phone, tablet, or desktop computer. GARP Learning allows an FRM candidate to engage meaningfully in a self-directed fashion with the full FRM Part I curriculum. It provides the ability to monitor performance, identify strengths and weaknesses, and assists in creating a personalized study plan.

Supplemental to the support offered by the learning platform, candidates can also utilize end-of-chapter questions to test their understanding of the chapter's content immediately; and, importantly, take a full-length FRM Part I Practice Exam to gain familiarity with how topics are tested and how to pace oneself on the exam to ensure completion in the allotted time.

As you can readily see, we are committed to ensuring the FRM program retains its global reputation as being of the highest quality, and covering the concepts, issues, and challenges that financial risk management professionals must know, and in many cases master.

As always, we wish you the very best as you study for the FRM exams, and much success in your career as a risk management professional.

Yours truly,

Richard Apostolik

President & CEO

FRM® COMMITTEE

Chairperson

Nick Strange, FCA (Chair)
Senior Technical Advisor, Operational Risk & Resilience,
Prudential Regulation Authority, Bank of England

Members

Richard Apostolik
President and CEO, GARP

Richard Brandt
Managing Director, Operational Risk Management, Citigroup

Julian Chen, FRM
SVP, FRM Program Manager, GARP

Chris Donohue
Managing Director, GARP Benchmarking Initiative, GARP

Donald Edgar, FRM
Managing Director, Risk & Quantitative Analysis, BlackRock

Hervé Geny
Former Group Head of Internal Audit, London Stock Exchange Group

Aparna Gupta, Ph.D.
Professor of Quantitative Finance
Associate Dean, Academic Affairs
A.W. Lawrence Professional Excellence Fellow
Co-Director and Site Director, NSF IUCRC CRAFT,
Lally School of Management
Rensselaer Polytechnic Institute

John Hull, Ph.D.
Senior Advisor
Maple Financial Professor of Derivatives and Risk Management,
Joseph L. Rotman School of Management, University of Toronto

Keith Isaac, FRM
VP, Capital Markets Risk Management, TD Bank Group

William May
Managing Director, Global Head of Certifications and Educational Programs, GARP

Attilio Meucci, Ph.D., CFA
Founder, ARPM

Victor Ng, Ph.D.
Chairman, Audit and Risk Committee
Former Managing Director, Head of Risk Architecture,
Goldman Sachs

Matthew Pritsker, Ph.D.
Senior Financial Economist and Policy Advisor/Supervision, Regulation, and Credit, Federal Reserve Bank of Boston

Samantha C. Roberts, Ph.D., FRM, SCR
Instructor and Consultant, Risk Modeling and Analytics

Til Schuermann, Ph.D.
Partner, Oliver Wyman

Evan Sekeris, Ph.D.
Head of Non-Financial Risk, MUFG

Sverrir Þorvaldsson, Ph.D., FRM
Senior Quant, SEB

ATTRIBUTIONS

Author

John C. Hull, PhD, University Professor, Maple Financial Group Chair in Derivatives and Risk Management, University of Toronto

Reviewers

Paul Feehan, PhD, Professor of Mathematics and Director of the Masters of Mathematical Finance Program, Rutgers University

Erick W. Rengifo, PhD, Associate Professor of Economics, Fordham University

Donald Edgar, FRM, MD, Risk & Quantitative Analysis, BlackRock

Jennifer Voitle, FRM, CFA, Senior Director of Model Risk Management Group, Santander Bank

Masao Matsuda, PhD, FRM, CAIA, President and CEO, Crossgates Investment and Risk Management

Tao Pang, PhD, FRM, CFA, Director of the Financial Mathematics Program, North Carolina State University

Donal McKillop, PhD, Professor of Financial Services, Queen's University in Belfast

Carlos Martin, FRM, Risk Director, Royal Bank of Scotland

Dennis Wong, FRM, CFA, PMP, Product Consultant, Thomson Reuters

Olivier Vu Liem, FRM, Principal, Société Générale

Sandeep Guhargarkar, FRM, Senior Director, Capital One

1

Measures of Financial Risk

Learning Objectives

After completing this reading, you should be able to:

- Describe the mean-variance framework and the efficient frontier.
- Compare the normal distribution with the typical distribution of returns of risky financial assets such as equities.
- Define the VaR measure of risk, describe assumptions about return distributions and holding periods, and explain the limitations of VaR.
- Explain and calculate ES and compare and contrast VaR and ES.
- Define the properties of a coherent risk measure and explain the meaning of each property.
- Explain why VaR is not a coherent risk measure.

Financial risk is measured in several ways. In this chapter, we start by discussing the mean and standard deviations of returns. Specifically, we show how the mean and standard deviation of a portfolio's return can be calculated using the means and standard deviations of its components' returns. This leads to an important concept known as the *efficient frontier*, which shows the trade-offs between mean and standard deviation that are available to the holder of a well-diversified portfolio.

When portfolios are described by the mean and standard deviation of their returns, it is natural to assume we are dealing with normal distributions. However, most financial variables have fatter tails than the normal distribution. In other words, extreme events are more likely to occur than the normal distribution would predict. This is relevant because risk managers are particularly interested in quantifying the probability of adverse extreme events.

Value-at-risk (VaR) and expected shortfall are two risk measures focusing on adverse events. Later in this chapter, we explain these measures and explore their properties. We find that expected shortfall, although less intuitive than VaR, has more desirable theoretical properties and is an example of a *coherent risk measure*.

1.1 THE MEAN-VARIANCE FRAMEWORK

Investors are faced with a trade-off between risk and return. The greater the risks that are taken, the higher the expected return that can be achieved. It should be emphasized that the term *expected return* does not describe the return that we expect to happen. Rather, the term is used by statisticians to describe the average (or mean) return.

One measure of risk is the standard deviation of returns. Consider a risk-free asset, such as a U.S. Treasury instrument. Suppose this asset provides a one-year return equal to 2%. The return, therefore, has a mean of 2% and a standard deviation of zero (when using a one-year horizon).

Consider next the situation presented in Table 1.1. The expected return is calculated by weighting the returns by their probabilities:

$$0.05 \times (-20\%) + 0.25 \times 0\% + 0.4 \times 7\% + 0.25 \times 15\% + 0.05 \times 40\% = 7.55\%$$

The return standard deviation is

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

Table 1.1 One-Year Returns and their Probabilities

Probability	Return
0.05	-20%
0.25	0%
0.4	7%
0.25	15%
0.05	40%

where E denotes expected value and R is the return. We have already calculated $E(R)$ as 7.55% (or 0.0755). The term $E(R^2)$ is calculated as follows:

$$0.05 \times (-0.2)^2 + 0.25 \times 0^2 + 0.4 \times 0.07^2 + 0.25 \times 0.15^2 + 0.05 \times 0.4^2 = 0.0176$$

The formula for the standard deviation gives

$$SD(R) = \sqrt{0.0176 - 0.0755^2} = 0.109$$

or 10.9%.

We have shown that the opportunity presented in Table 1.1 has a return with a mean of 7.55% and a standard deviation of 10.9%. We can then characterize the set of all investment opportunities in a chart by plotting the various combinations of means and standard deviations. Figure 1.1 shows the two assets we have mentioned so far:

1. An asset having a return with a mean of 2.00% and a standard deviation of zero
2. An asset having a return with a mean of 7.55% and a standard deviation of 10.9%

Combining Investments

Now consider two available investments with expected returns μ_1 and μ_2. Suppose an investor puts a proportion w_1 of available

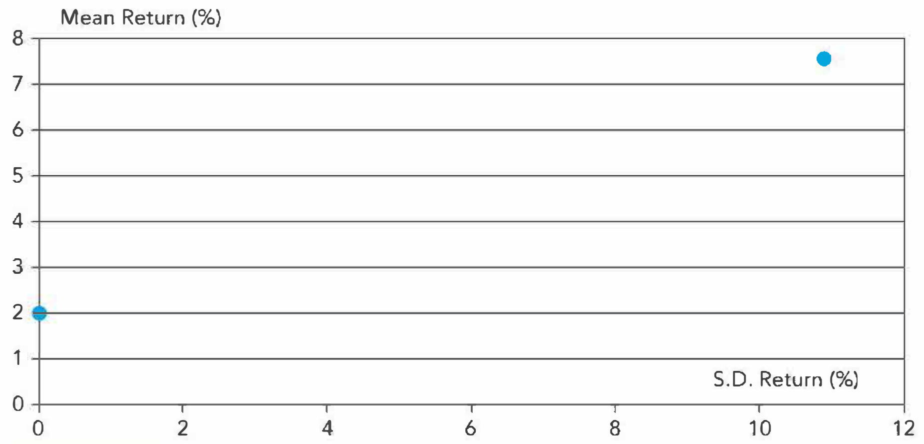


Figure 1.1 Two alternative risk-return possibilities.

Table 1.2 Risk-Return Combinations from Two Investments

w_1	w_2	μ_P	σ_P
0.0	1.0	8.0%	16.0%
0.2	0.8	7.4%	13.6%
0.4	0.6	6.8%	11.8%
0.6	0.4	6.2%	10.8%
0.8	0.2	5.6%	10.9%
1.0	0.0	5.0%	12.0%

funds into investment 1 and $w_2 = 1 - w_1$ into investment 2. The portfolio expected return (μ_P) is then a weighted average of the expected return from the two investments:

$$\mu_P = w_1\mu_1 + w_2\mu_2 \tag{1.1}$$

The standard deviation of the portfolio's return (σ_P) is more complicated and involves the coefficient of correlation between the returns from the two investments (ρ):

$$\sigma_P = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2\rho w_1 w_2 \sigma_1 \sigma_2} \tag{1.2}$$

Suppose that $\mu_1 = 5\%$ and $\mu_2 = 8\%$; $\sigma_1 = 12\%$ and $\sigma_2 = 16\%$; and $\rho = 0.25$. Table 1.2 shows several possible outcomes for various weightings for investment 1 and investment 2. These are plotted in Figure 1.2.

We can extend the previous result to the situation in which there are n investments and μ_i and σ_i are respectively the mean and standard deviation of the return for the ith investment. When the weight given to the ith investment is w_i, the mean and standard deviation of the portfolio return are

$$\mu_P = \sum_{i=1}^{n} w_i\mu_i \tag{1.3}$$

and

$$\sigma_P = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} \rho_{ij} w_i w_j \sigma_i \sigma_j} \tag{1.4}$$

where ρ_{ij} is the correlation between the returns from investments i and j.

Investors want high expected returns and low risks. If we measure risk by the standard deviation of the return, we can argue that some portfolios in Table 1.2 and Figure 1.2 dominate others for these investors. For example, given a choice between the $w_1 = 1.0$ (where $\mu_p = 5.0\%$ and $\sigma_p = 12.0\%$) and $w_1 = 0.6$ (where $\mu_p = 6.2\%$ and $\sigma_p = 10.8\%$), investors would seem to prefer the latter because its return has a higher mean (6.2% versus 5.0%) and a lower standard deviation (10.8% versus 12.0%).

Of the six investments in Figure 1.2, the upper four (corresponding to $w_1 = 0.0$, 0.2, 0.4, and 0.6) form what is referred to as an efficient frontier. For each of these four investments, there is no other investment (or combination of investments) that has both a higher expected return and a lower standard deviation of return.

The Efficient Frontier

As mentioned, most investors want to maximize their expected returns and minimize their risk. If we measure risk by the return standard deviation, we can surmise that these investors want to move as far into the upper-left region in Figure 1.2 as possible.

Using Equations (1.3) and (1.4), we find that an investor can move further in the desired direction by incorporating additional investments. For example, a third investment can be combined with any combination of the first two investments to obtain new risk-return trade-offs. A fourth investment can be combined with any combination of the first three investments to obtain even more risk-return trade-offs, and so on.

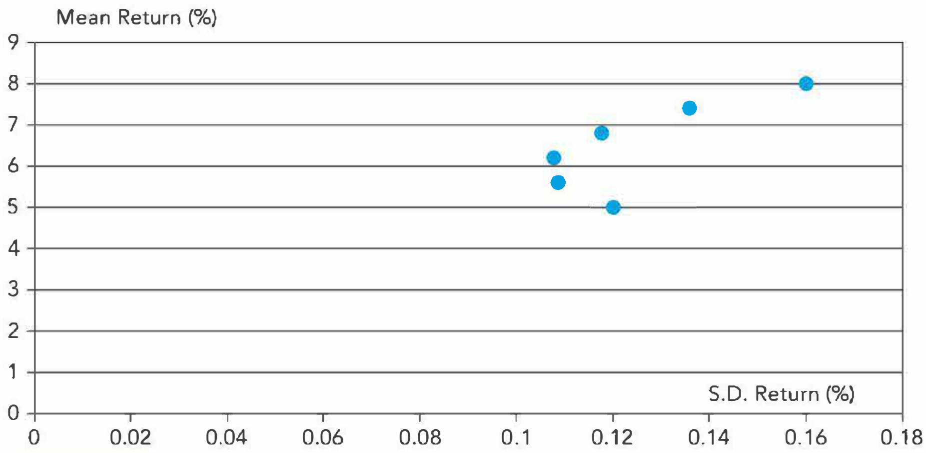


Figure 1.2 Risk-return combinations in Table 1.2.

Of course, there are limits as to how favorable the risk-return combinations can become. When all possible combinations of all risky investments are considered, we obtain the efficient frontier of all risky investments. This is shown in Figure 1.3. All risk-return combinations in the shaded area below and to the right of the efficient frontier can be created. Those to the upper-left cannot be created.

So far, we have only considered risky investments in constructing the efficient frontier. We now consider what happens when a risk-free investment earning a return of R_F is

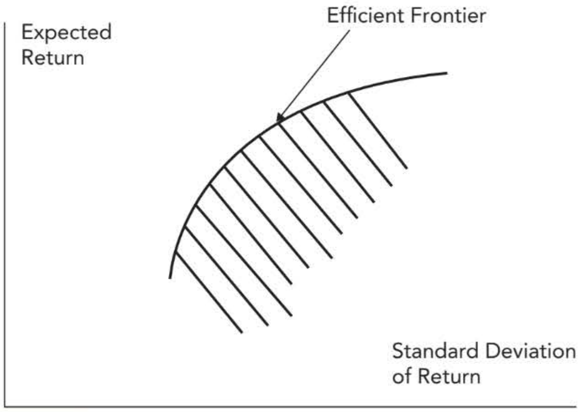


Figure 1.3 Efficient frontier obtained from risky investments.

possible. This investment is represented by point F in Figure 1.4. Consider a line drawn from F that is a tangent to the efficient frontier in Figure 1.3. Suppose that M is the point of tangency. By dividing our funds between the risk-free investment represented by point F and the risky portfolio represented by point M, we can obtain the risk-return combinations represented by any point on the line FM. For example, the risk-return combination represented by the mid-point of FM is obtained by putting half of the available funds in the risk-free asset and the other half in the risky portfolio represented by point M. The portfolio represented by the point that is a quarter of the way from F to M can be created by putting 75% of available funds in the risk-free asset and 25% in the portfolio represented by point M.

We can prove this observation using Equations (1.1) and (1.2). Suppose investment 1 is the risk-free asset and investment 2 is the portfolio represented by point M. For the risk-free asset, the expected return is R_F and the standard deviation of the return

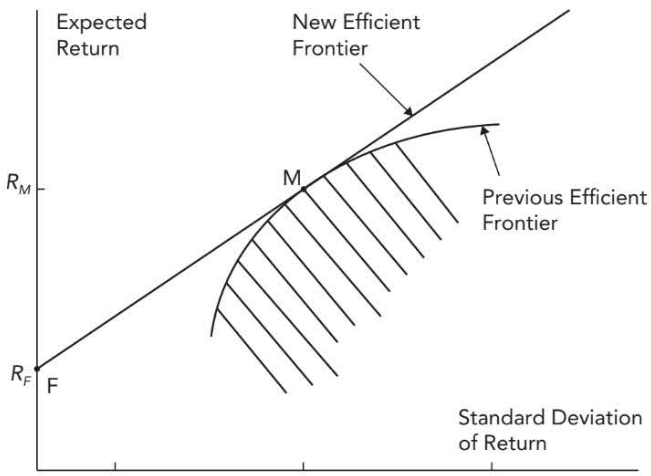


Figure 1.4 Efficient frontier when borrowing and lending at the risk-free rate is possible.

is zero. Suppose that, as indicated in Figure 1.4, the expected return and standard deviation of the return for the portfolio represented by point M are μ_M and σ_M. If we put a proportion β in the portfolio represented by point M and a proportion $1-\beta$ in the risk-free asset, we can use Equations (1.1) and (1.2) with $w_1 = 1-\beta$, $w_2 = \beta$, $\mu_1 = R_F$, $\mu_2 = R_M$, $\sigma_1 = 0$, and $\sigma_2 = \sigma_M$ to show that the expected return is

$$(1-\beta)R_F + \beta R_M$$

and the standard deviation of the return is $\beta\sigma_M$. When $\beta = 0.5$ this corresponds to the point half way between F and M; when $\beta = 0.25$, it corresponds to the point quarter of the way from F to M, and so on.

We can carry this analysis further. If we make the assumption that we can borrow at R_F, we can obtain points on the line FM that are beyond M. Suppose, for example, that $\beta > 1$ and that we borrow USD ($\beta-1$) for every USD 1 that we have available for investment. The total funds available for investment is then β multiplied by the original funds. Assume that it is all invested in the portfolio represented by point M. When the interest on the borrowings is considered, the return is

$$\beta R_M - (\beta - 1)R_F = (1 - \beta)R_F + \beta R_M$$

The standard deviation of the return is $\beta\sigma_M$. As a function of β, these are the same as before and the risk-return combinations are those represented by the points on the continuation of line FM beyond M.

This argument shows that, when a risk-free investment is considered and borrowing (as well as lending) at the risk-free rate is possible, the efficient frontier must be a straight line. There is a linear trade-off between expected return and standard deviation of return. The argument also shows that all investors should choose to invest in the same portfolio of risky assets, for example, the portfolio represented by point M. They should then reflect their risk appetite by borrowing or lending at the risk-free rate, R_F. Investors who are relatively risk-averse will choose points on the line FM that are close to F. Those who like taking risks will choose points close to M, or even points on the line FM that are beyond M.

The investment represented by point M is referred to as the market portfolio. It is the portfolio consisting of all investments in the market with the proportional amount of any given investment in the portfolio being the same as the proportion of all available investments that it represents. To understand why this must be so, first consider the situation where investment X is under-represented in portfolio M. Because all investors want to buy portfolio M, demand for investment X will be less than supply, and its price will decline to a point where it is no longer under-represented in M. Similarly, consider the opposite situation where investment X is over-represented in portfolio M. In this case, demand for investment X would exceed supply, and its price would increase until the over-representation disappeared.

We have tacitly assumed that all investors make the same assumptions about the mean and standard deviations of, and coefficients of correlation between, the returns from different investments. We have also assumed that they care only about the mean and standard deviation of the returns from their portfolios, and that they can all borrow at the risk-free rate of interest. These assumptions are, at best, only approximately true.

Efficient frontiers are closely related to the capital asset pricing model (CAPM). This is discussed further in the Foundations of Risk Management readings.

1.2 THE NORMAL DISTRIBUTION

The normal (or Gaussian) distribution is one of the most well-known and widely used probability distributions. It has two parameters: the mean and standard deviation. A normally distributed variable x, with mean μ and standard deviation σ, has a density function:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1.5}$$

Figure 1.5 shows the density function for $\mu = 2$ and $\sigma = 5$. The center of the distribution is μ, and σ is a measure of dispersion about the mean.

The height of the normal distribution for a value of x can be thought of as a measure of the probability of that value of x (or a value close to it) occurring. Values close to the center of the distribution are most likely to occur, whereas values at the tails of the distribution are less likely to occur.

As with all probability density functions, the probability of a value between a and b occurring is equal to the area under the probability density function between a and b. The cumulative density function provides the area under a distribution up to a certain point. The probability of a value between a and b is equal to the cumulative distribution up to b minus the cumulative density function up to a.

A standard normal distribution is one where the mean is zero and the standard deviation is 1 so that the density function is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{1.6}$$

Tables are often used to provide the cumulative distribution for the standard normal distribution. For example, the area between minus infinity and 1.0 is 0.8413. For a distribution with mean μ and standard deviation σ, we can use z-scores. The cumulative probability up to a value x is the cumulative probability for a standard normal up to

$$z = \frac{x - \mu}{\sigma}$$

For example, when $\mu = 2$ and $\sigma = 5$, the probability that the value is less than 1 is the probability that a standard normal variable is less than

$$\frac{1 - 2}{5} = -0.2$$

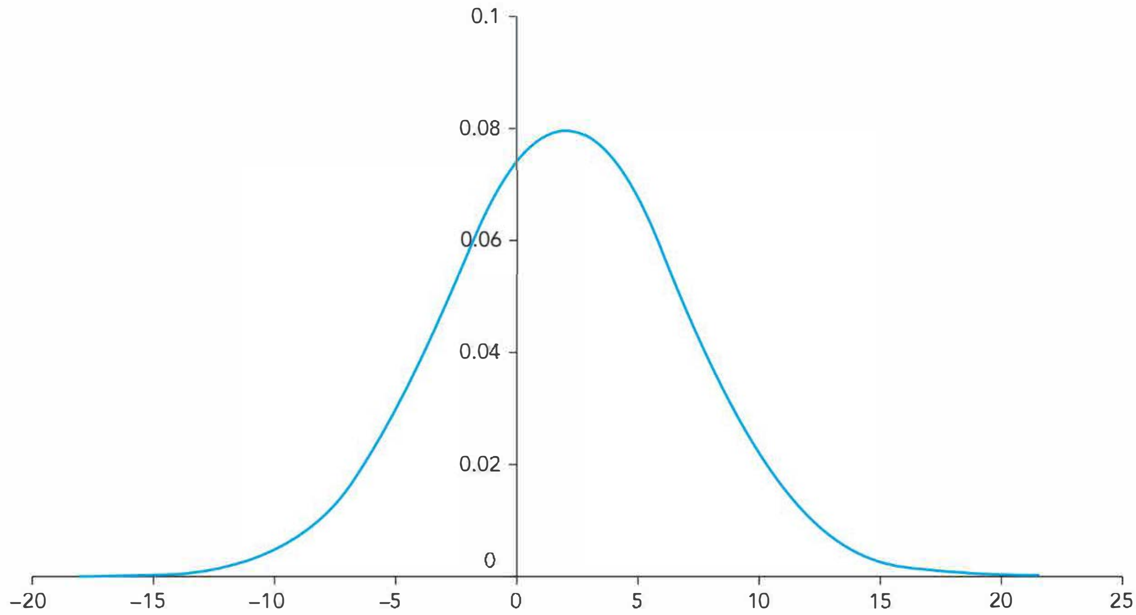


Figure 1.5 **A normal distribution with mean 2 and standard deviation 5.**

Table 1.3 Percentage of Days when the Change in the S&P 500 is Greater than a Number of Standard Deviations and the Predictions Made by the Normal Model

Movement	Actual Results (%)	Predicted by Normal Distribution (%)
>1 SD	21.05	31.73
>2 SD	5.13	4.55
>3 SD	1.64	0.27
>4 SD	0.64	0.01
>5 SD	0.30	0.00
>6 SD	0.12	0.00

From tables, this is 0.4207.

Since 2007, Excel has provided a convenient function for the normal distribution. For example, NORM.DIST(1, 2, 5, TRUE) gives the probability that a normal variable with mean 2 and standard deviation 5 is less than 1. As we have just mentioned, that probability is 0.4207 Similarly, NORM.DIST(4, 2, 5, TRUE) gives the probability of a value less than 4. This is 0.6554. The probability of a value between 1 and 4 is 0.2347 (=0.6554 − 0.4207).

The normal distribution is often assumed for the returns on financial variables. It is a convenient choice because all we need is the mean and standard deviation of the returns, and these can be calculated for a portfolio from Equations (1.3) and (1.4).

In practice, financial variables tend to have much fatter tails than normal distributions. We will illustrate this using data on the S&P 500 Index. Over a 20-year period between January 1, 2000, and December 31, 2019, the standard deviation of the daily change in the index was 1.187%.[1] Table 1.3 shows

- The percentage of days in which the actual daily return exceeded one, two, three, four, five, and six standard deviations; and
- The percentage of days we would expect this to happen if the probability of daily returns had been normally distributed.

Consider, for example, the row in Table 1.3 corresponding to ">3 SD." The daily change in the S&P 500 was greater than three standard deviations (or 3 × 1.187% = 3.56%) on 82 days, or 1.64% of the time. A normal probability distribution predicts that a variable will exceed three standard deviations only 0.27% of the time.[2] A three standard deviation move is therefore over six times more likely than the normal distribution would predict. Looking at other rows in Table 1.3, we see that the normal distribution predicts that four, five, and six standard deviation moves should hardly ever happen. In practice, we observe them on 0.64%, 0.30%, and 0.12% of the days, respectively.

Most financial variables are like the S&P 500 in that they have fatter tails than the normal distribution. Most portfolio returns that can be created also have fatter tails than the normal distribution. The situation is illustrated in Figure 1.6. The actual distribution and the normal distribution have the same standard deviation. However, the actual distribution is more peaked than the normal distribution and has fatter tails. This means that both small changes and large changes happen more often than the normal distribution would suggest, while intermediate changes happen less often. This is consistent with our S&P 500 data. The probability of a percentage change less than one standard deviation is 100% − 21.05% = 78.95%, whereas the normal distribution would predict it to be 100% − 31.73% or 68.27%. The probability of a percentage change greater than three standard deviations is 1.64% compared with 0.27% for the normal distribution. The probability of a percentage change between one and three standard deviations is 21.05% − 1.64% = 19.41%, whereas the probability of this for the normal distribution is 31.73% − 0.27% = 31.46%.

1.3 VaR

Risk managers are concerned with the possibility of adverse events. We have shown that describing a probability distribution by its mean and standard deviation and then assuming a normal distribution is likely to underestimate the probability of extreme events. The standard deviation of a distribution, although a useful measure in many situations, does not describe the tails of a probability distribution. VaR is an important risk measure that focuses on adverse events and their probability.

The VaR for an investment opportunity is a function of two parameters:

- The time horizon, and
- The confidence level.

[1] For the purposes of this example, we assume that the mean daily return on the S&P 500 was zero. Over short time periods, such as one day, the standard deviation of the return is much more important than the mean return. For our data, the mean daily return was only 0.024%, so that the daily standard deviation was almost 50 times the mean.

[2] The Excel function NORM.DIST(-3) gives 0.00135 indicating that a decline of three standard deviations or more has a probability of 0.135%. Similarly, the probability of an increase of three standard deviations or more is 0.135%. The probability of a movement exceeding three standard deviations is therefore 2 × 0.135% or 0.27%.

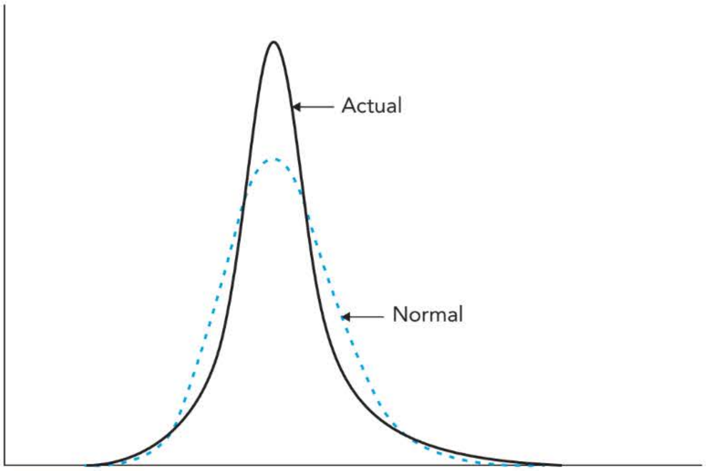


Figure 1.6 Normal distribution compared with a typical fat-tailed distribution that is often encountered in practice.

VaR is the loss level that we do not expect to be exceeded over the time horizon at the specified confidence level.

Whereas in the previous section we considered the distribution of returns, in the remainder of this chapter we are concerned with the distribution of the dollar amount of losses. Gains are counted as negative losses.

Suppose that the time horizon is ten days, and the confidence level is 99%. A VaR of USD 10 million would mean that we are 99% certain that the loss during the next ten days will be less than USD 10 million. To put this another way, there is a probability of only 1% that the loss over the next ten days will be greater than USD 10 million.

Suppose that the probability density function for the loss during the time horizon is as shown in Figure 1.7. The VaR with a confidence level of X% is found by searching for the loss that has an $(100 - X)$% chance of being exceeded. As indicated in Figure 1.7, it is the value such that the area under the distribution equals $(100 - X)$%.

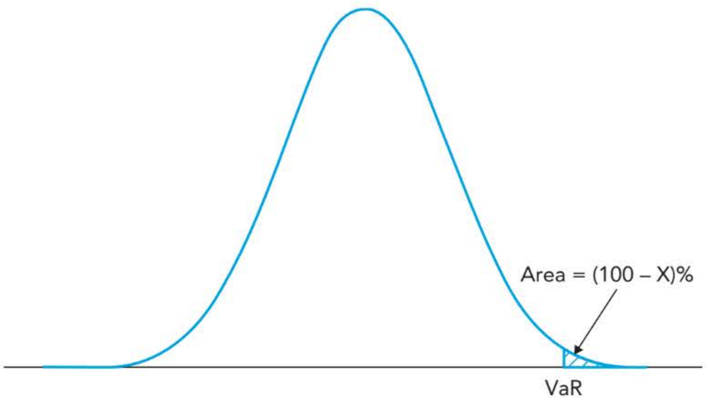


Figure 1.7 Calculation of VaR from loss distribution when the confidence level is *X*%.

As a simple example of VaR, suppose the return from an investment over a specified time horizon has a normal distribution with mean 20 and standard deviation 30. This corresponds to a loss distribution with mean −20 and standard deviation 30. Suppose that we wish to calculate the VaR with a confidence level of 99%. This is the loss level that has a 1% chance of being exceeded. The function NORM.INV in Excel can be used for this calculation. The first argument is the percentile of the distribution required (0.99), the second is the mean loss (−20), and the third is the standard deviation (30). NORM.INV(0.99,−20,30) gives 49.79. Thus, the 99th percentile of a normal loss distribution with a mean of −20 and a standard deviation of 30 is 49.79. This is the VaR level when the confidence level is 99%.

As a second simple example, assume the result of an investment with a uniform distribution where all outcomes between a profit of 30 and a loss of 20 are equally likely. In this case, the VaR with a 99% confidence level is 19.5. This is because the probability that the loss will lie between 19.5 and 20 is $0.5/50 = 1\%$. (Note that we divide the range of losses we are interested in (0.5) by the total range of losses (50) because all outcomes are equally likely.)

As a final example, let us suppose that the outcomes are discrete rather than continuous. Suppose that a project with one year remaining is performing badly and is certain to lead to a loss. Three different scenarios are possible:

1. A loss of USD 2 million (probability 88%),
2. A loss of USD 5 million (probability 10%), and
3. A loss of USD 8 million (probability 2%).

This is summarized in Table 1.4. The final column shows the cumulative loss probability measured from the lowest to the highest. The first 88% of the cumulative loss distribution corresponds to a loss of USD 2 million. The part of the cumulative distribution between 88% and 98% corresponds to a loss of USD 5 million. The part of the cumulative distribution between 98% and 100% corresponds to a loss of USD 8 million.

With a 99% confidence level, the VaR is USD 8 million. This is because 99% falls into the 98% to 100% cumulative probability range, and the loss for this range is USD 8 million. When the

Table 1.4 Discrete Loss Example

Loss (million)	Probability (%)	Cumulative Probability Range (%)
2	88	0 to 88
5	10	88 to 98
8	2	98 to 100

confidence level is reduced to 97%, the VaR is USD 5 million because 97% falls into the 88% to 98% cumulative probability range, and the loss for this range is USD 5 million.

Note that if the confidence level is 98%, there is some ambiguity. We could argue that we are in the 98% to 100% range so that the VaR is USD 8 million, or that we are in the 88% to 98% range so that the VaR is USD 5 million. One approach here is to set the VaR equal to the average of the two answers, or USD 6.5 million.

1.4 EXPECTED SHORTFALL

One problem with VaR is that it does not say anything about how bad losses might be when they exceed the VaR level. Suppose that the VaR with a 99% confidence level is USD 20 million. We can imagine two situations.

1. If the loss is greater than USD 20 million (a 1% chance), it is inconceivable that it will be greater than USD 30 million.
2. If the loss is greater than USD 20 million (a 1% chance), it is most likely to be USD 100 million.

The VaR for both situations is USD 20 million, but the second situation is clearly riskier than the first. The second situation could arise when a trader has sold a credit default swap as part of a portfolio. This is a contract that provides a payoff if a particular counterparty defaults. The default might have a small probability, say 0.9%, but lead to a loss of USD 100 million. VaR does not distinguish between the two situations because it sets the risk measure equal to a particular percentile of the loss distribution and takes no account of possible losses beyond the VaR level.

Expected shortfall is a risk measure that does take account of expected losses beyond the VaR level. Expected shortfall, which is also called conditional VaR (C-VaR) or tail loss, is the expected loss conditional that the loss is greater than the VaR level. In the example given previously, the expected shortfall with a 99% confidence level would be much greater in the second situation (when the loss is greater than USD 20 million, it is most likely to be USD 100 million) than in the first situation (where losses in excess of USD 20 million will not exceed USD 30 million).

When losses are normally distributed with mean μ and standard deviation σ, the expected shortfall is

$$\mu + \sigma \frac{e^{-U^2/2}}{(1 - X)\sqrt{2\pi}} \tag{1.7}$$

where X is the confidence level, and U is the point in the standard normal distribution that has a $(100 - X)\%$ probability of being exceeded. In an earlier example, the loss was normally distributed with mean -20 and standard deviation 30. The VaR with a 99% confidence level was 49.79. When $X = 99\%$, $U = 2.326$ and the expected shortfall, given by equation (1.7), is

$$-20 + 30 \frac{e^{-2.326^2/2}}{0.01 \times \sqrt{2 \times 3.1416}} = 59.96$$

Note that expected shortfall must, by definition, be greater than VaR.

Now consider again the example where all outcomes between a profit of 30 and a loss of 20 are equally likely. As explained above, the VaR with 99% confidence is 19.5 because losses in excess of 19.5 (i.e., between 19.5 and 20) have a probability of 1%. Conditional on the loss being greater than 19.5, the expected loss is half way between 19.5 and 20, or 19.75. The expected shortfall with a confidence level of 99% is therefore 19.75.

Finally, let us return to the example in Table 1.4 where losses of USD 8 million, USD 5 million, and USD 2 million have probabilities of 2%, 10%, and 88% respectively. The VaR with a 97% confidence is USD 5 million. To calculate expected shortfall, we must ask ourselves, "If we are in the worst 3% of the loss distribution, what is the expected loss?" The final column of Table 1.4 makes it clear that the 3% tail of the distribution is composed of a 2% probability that the loss is USD 8 million and a 1% probability that the loss is USD 5 million. Conditional on being in the tail of the distribution, there is therefore a 2/3 chance that the loss is USD 8 million and a 1/3 chance that it is USD 5 million. The expected shortfall (in millions of dollars) is therefore:

$$(2/3) \times 8 + (1/3) \times 5 = 7$$

1.5 COHERENT RISK MEASURES

Regulators have historically used VaR to determine the capital banks must keep when internal models are used. For example, the rules introduced for market risk in 1996 based the required capital on the VaR with a ten-day time horizon and a 99% confidence level. Meanwhile, Basel II based credit risk capital on the VaR with a one-year time horizon and a 99.9% confidence level. Regulators currently intend to bring in a new way of calculating market risk capital known as FRTB. This will replace VaR with expected shortfall and lower the confidence level to 97.5%.

What is the best risk measure to use when determining capital requirements? Artzner et al. have examined this question theoretically.[3] They first proposed four properties a risk measure should have. These are as follows.

[3] See P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Coherent Measures of Risk," Mathematical Finance 9 (1999): 203–228.

1. *Monotonicity:* If (regardless of what happens) a portfolio always produces a worse result than another portfolio, it should have a higher risk measure.
2. *Translation Invariance:* If an amount of cash K is added to a portfolio, its risk measure should decrease by K.
3. *Homogeneity:* Changing the size of a portfolio by multiplying the amounts of all the components by λ results in the risk measure being multiplied by λ.
4. *Subadditivity:* For any two portfolios, A and B, the risk measure for the portfolio formed by merging A and B should be no greater than the sum of the risk measures for portfolios A and B.

A risk measure that satisfies all four conditions is termed *coherent*.

These conditions are fairly intuitive. The first property simply says that if one portfolio always performs worse than another, it is riskier (i.e., requires more capital). The second condition seems reasonable because cash provides a cushion against losses and can be regarded as a substitute for capital. Meanwhile, if we halve the size of a portfolio, its risk measure should also be cut in half as specified by the third condition.[4] Finally, the fourth property simply reflects the benefits of diversification. If A and B are perfectly correlated, the total risk should be the sum of the risks considered separately. If A and B are less than perfectly correlated, there are diversification benefits, and we should expect the risk measure to decrease.

Artzner et al. showed that expected shortfall has all four properties, while VaR has only the first three properties. There are circumstances when VaR does not satisfy the fourth diversification property.

We can illustrate Artzner et al.'s result with a simple example. Suppose two insurance contracts each have a probability of 0.8% of a loss of USD 20 million and a probability of 99.2% of a loss of USD 1 million. The VaR with 99% confidence for each contract is USD 1 million. When we are in the 1% tail of the loss distribution there is an 80% chance that the loss will be USD 20 million and a 20% chance that it will be USD 1 million. The expected shortfall of each contract is therefore $(0.8 \times 20 + 0.2 \times 1)$ million or USD 16.2 million.

Now suppose that the two contracts are independent. There is a $0.008 \times 0.008 = 0.000064$ probability that both contracts will lose USD 20 million for a total loss of USD 40 million. There is a $0.008 \times 0.992 = 0.007936$ probability that the first contract will lose USD 20 million, and the second one will lose USD 1 million for a total loss of USD 21 million. Similarly, there is a 0.007936 probability that the first contract will lose USD 1 million, and the second will lose USD 20 million for a total loss of USD 21 million. There is, therefore, a total probability of $2 \times 0.007936 = 0.015872$ of a loss of USD 21 million. Finally, there is a $0.992 \times 0.992 = 0.984064$ chance that each contract will lose USD 1 million for a total loss of USD 2 million. The outcomes from the two contracts are as indicated in Table 1.5.

Table 1.5 **Losses for Two-Contract Example**

Loss (USD millions)	Probability (%)	Cumulative Probability Range (%)
2	98.4064	0 to 98.4064
21	1.5872	98.4064 to 99.9936
40	0.0064	99.9936 to 100

The VaR for a portfolio of the two contracts is the loss that corresponds to the 99% point on the cumulative loss distribution. This is USD 21 million (because 99% falls within the 98.4064% to 99.9936% range). To calculate the expected shortfall, we note for outcomes in the region between the 99% and 100% point of the cumulative distribution a loss of USD 40 million has a probability of 0.000064, and a loss of USD 21 million has a probability of 0.009936. The expected shortfall in USD million is therefore:

$$\frac{0.000064 \times 40 + 0.009936 \times 21}{0.01} = 21.1236$$

Expected shortfall satisfies the fourth subadditivity condition because each contact taken on its own has an expected shortfall of USD 16.2 million, and the two contracts taken together have an expected shortfall of USD 21.1236 million with:

$$21.1236 < 16.2 + 16.2$$

However, VaR does not satisfy the subadditivity condition. Each contract considered on its own has a VaR of USD 1 million. The two contracts together have a VaR of USD 21 million. Combining the two projects leads to an increase of USD 19 million in the total VaR.

Weighting and Spectral Risk Measures

Suppose you have calculated a probability distribution for the outcomes in a certain situation. A risk measure can be defined by assigning weights to the percentiles of the loss distribution. Consider the two risk measures we have introduced from the perspective of how weights are assigned.

[4] Note that liquidity considerations may lead to the third condition being untrue. If a portfolio is multiplied by 100 it may be more than 100 times as risky because of the difficulty of unwinding it.

VaR with a confidence level of $X\%$ is simply the Xth percentile of the loss distribution. All the weight is assigned to one specific percentile. As pointed out earlier, the measure takes no account of losses more than the X percentile point. We can consider two investments, A and B, where in each case a loss level of V has a 1% probability of being exceeded. In A, losses will never exceed $1.1V$. In B, losses as great as $10V$ are possible. Both A and B have a VaR equal to V when the confidence level is 99%, but B is clearly riskier than A.

Expected shortfall assigns a probability-based weight to all loss levels greater than the VaR. In the example we have just given, project B would give rise to a much higher expected shortfall than project A.

It turns out that a risk measure is coherent, as defined previously, if and only if the weights are a non-decreasing function of the percentile of the loss distribution. Suppose that $w(p)$ is the weight assigned to the p-percentile. For a coherent risk measure:

$$w(p_1) \geq w(p_2) \text{ when } p_1 > p_2$$

VaR does not satisfy this condition because all the weight is given to one particular percentile. Suppose that the confidence level is 99%. All the weight is assigned to the 99th percentile. This means that if we put p_2 equal to 99% and p_1 equal to, say, 99.5% we find that $w(p_2) = 1$ while $w(p_1) = 0$, which violates the condition.

Expected shortfall does satisfy the condition. The weights have the pattern shown by the solid line in Figure 1.8. No weight is given to losses that are less than the VaR loss level. All losses greater than the VaR loss level are given the same weight. The weights are therefore always a non-decreasing function of the percentile.

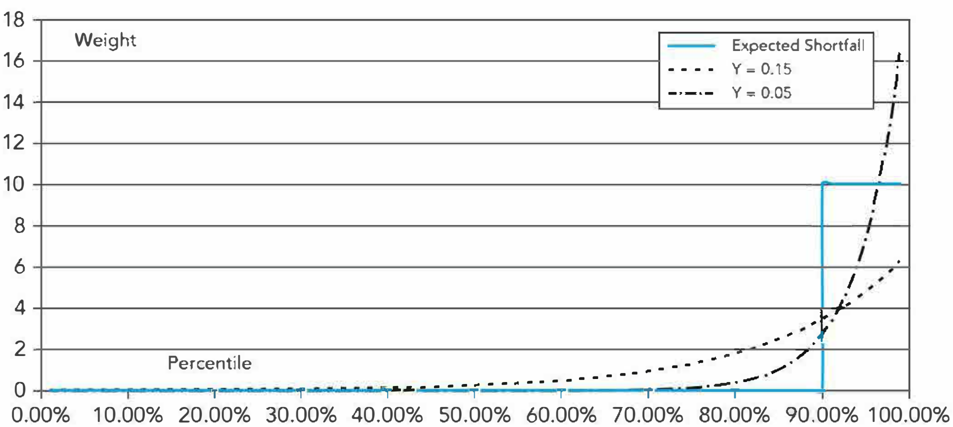


Figure 1.8 **Weights as a function of percentiles for risk measures.**

Expected shortfall is by no means the only coherent risk measure than can be constructed. For example, a spectral risk measure is a type of coherent risk measure where the weights assigned to percentiles increase in a way that reflects risk aversion. One way that has been suggested to achieve this is to make weights proportional to:

$$e^{-(1-p)/\gamma}$$

where p is the percentile (expressed in decimal form), and γ is a constant reflecting the user's degree of risk aversion. The lower the value of γ, the greater is the degree of risk aversion. Figure 1.8 shows the variation of weights with percentiles for values of γ equal to 0.05 and 0.15, respectively.

SUMMARY

Risk measures are important in risk management. Standard deviation has been used as a risk measure for many years and is a good way of describing the overall uncertainty associated with a set of outcomes. When two or more investments are combined, it is easy to calculate the standard deviation of the entire portfolio return using the returns on the individual investments and the correlations between them.

Sometimes risk managers are comfortable assuming that the outcome from an investment (measured either as a return or in dollar terms) is normally distributed. An estimated mean and standard deviation are then sufficient to characterize outcomes. Usually, however, risk managers are concerned with carefully quantifying outcomes where losses are high. Often the returns from financial investments have fatter tails than the normal distribution. (We illustrated this using daily data over a 20-year period for the S&P 500.) This has led risk managers to search for ways of quantifying the tails of loss distributions.

VaR is the loss over a certain period that is not expected to be exceeded with a certain confidence level $X\%$. It is the Xth percentile point of the loss distribution. A key disadvantage of VaR is that it does not consider the magnitude of losses beyond the VaR level. Expected shortfall is one way of overcoming this disadvantage. It is defined as the expected (or average) loss conditional on the loss being greater than the VaR level.

A coherent risk measure is one that always satisfies four reasonable axioms (i.e., monotonicity, translation invariance, homogeneity, and subadditivity). It can be shown that expected shortfall is coherent whereas VaR is not. The axiom that is not always satisfied by VaR is subadditivity, which requires that a risk measure of a combination of two or more portfolios never be greater than the sum of the risk measures of each portfolio component. This axiom is sometimes violated for VaR, but never for expected shortfall.

It can be shown that a risk measure is coherent only if the weight assigned to a loss percentile is a non-decreasing function of the percentile. This allows coherent risk measures other than expected shortfall to be defined. Spectral risk measures are coherent risk measures where the weights assigned to percentiles reflect risk aversion.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

QUESTIONS

Short Concept Questions

1.1 How is the expected value of a variable defined?

1.2 What is a coherent risk measure?

1.3 What are the properties of a portfolio that is represented by a point on the efficient frontier?

1.4 How does the distribution of the daily return on the S&P 500 differ from a normal distribution?

1.5 How is VaR defined?

1.6 How is expected shortfall defined?

1.7 In what way does expected shortfall have better theoretical properties than VaR?

1.8 Explain the homogeneity and subadditivity conditions necessary for a coherent risk measure.

1.9 When a risk measure is defined by assigning weights to percentiles, under what circumstances is it coherent?

1.10 Under what circumstances is a spectral risk measure coherent?

Practice Questions

1.11 An investment has probabilities of 0.1, 0.3, 0.2, 0.3, and 0.1 of giving one-year returns equal to 30%, 20%, 10%, 0%, and −10%. What is the mean return and the standard deviation of the return?

1.12 Suppose that there are two investments with the same probability distribution of returns as in Question 11. What is the total mean and standard deviation of returns if the correlation between them is 0.2?

1.13 For the two investments considered in Table 1.2 and Figure 1.2, what are the risk-return combinations if the correlation is 0.15 instead of 0.25?

1.14 The argument in this chapter leads to the surprising conclusion that all investors should choose the same risky portfolio. What assumptions are necessary for this result?

1.15 The distribution of the losses from a project over one year has a normal loss distribution with a mean of −10 and a standard deviation of 20. What is the one-year VaR when the confidence level is (a) 95%, (b) 99%, and (c) 99.9%?

1.16 The change in the value of a portfolio over one day is normally distributed with a mean of 0.5 and a standard deviation of 2. What are the mean and standard deviation over ten days? Assume that changes in successive days are independent.

1.17 An investment has a uniform distribution where all outcomes between −40 and +60 are equally likely. What are the VaR and expected shortfall with a confidence level of 95%?

1.18 A one-year project has a 3% chance of losing USD 10 million, a 7% chance of losing USD 3 million, and a 90% chance of gaining USD 1 million. What are (a) the VaR and (b) the expected shortfall when the confidence level is 95% and the time horizon is one year?

1.19 Suppose that there are two independent identical investments with the properties specified in question 18. What are (a) the VaR and (b) the expected shortfall for a portfolio consisting of the two investments when the confidence level is 95% and the time horizon is one year?

1.20 Check whether (a) VaR or (b) expected shortfall satisfy the subadditivity axiom for a coherent risk measure for the investments in Questions 18 and 19.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

ANSWERS

Short Concept Questions

1.1 The expected value of a variable is its average (or mean) value.

1.2 A coherent risk measure is a risk measure that has satisfied four specified axioms. (See Section 1.5)

1.3 It is a portfolio such that there is no other portfolio with both a higher expected return and a lower standard deviation of return.

1.4 The distribution of the S&P 500 has fatter tails and is more peaked than the normal distribution.

1.5 VaR is the loss that will not be exceeded over a certain period with a certain probability. The probability is referred to as the confidence level.

1.6 The expected shortfall is the expected loss conditional on the loss being greater than the VaR level.

1.7 Expected shortfall satisfies the subadditivity (diversification) axiom. The expected shortfall from two portfolios is never more than the sum of the expected shortfall of the individual portfolios (when the same time horizon and confidence level is used throughout).

1.8 The homogeneity axiom states that changing the size of a portfolio by multiplying all its components by λ results in the risk measure being multiplied by λ. The subadditivity axiom states that for any two portfolios, A and B, the risk measure for the portfolio formed by merging A and B should be not greater than the sum of the risk measures for portfolios A and B.

1.9 Weights must be a non-decreasing function of the percentile.

1.10 A spectral risk measure is always coherent.

Solved Problems

1.11 The mean return is 10%. The expected squared return is 0.024 so that the standard deviation of the return is $\sqrt{0.024 - 0.1^2} = \sqrt{0.014} = 0.118$ or 11.8%.

1.12 The total mean return is 10%. The standard deviation of returns is

$$\sqrt{0.5^2 \times 0.118^2 + 0.5^2 \times 0.118^2 + 2 \times 0.5 \times 0.5 \times 0.118 \times 0.118 \times 0.2}$$

$= 0.0914$ or 9.14%.

1.13 The table becomes

w_1	w_2	μ_P	σ_P
0.0	1.0	8.0%	16.0%
0.2	0.8	7.4%	13.4%
0.4	0.6	6.8%	11.4%
0.6	0.4	6.2%	10.3%
0.8	0.2	5.6%	10.6%
1.0	0.0	5.0%	12.0%

1.14 There are several assumptions, for example:

- Investors care only about mean return and standard deviation of return.
- Investors can borrow and lend at the same risk-free rate.
- All investors make the same estimates of expected return and standard deviation of return.
- There are no tax considerations.

1.15 (a) 22.9, (b) 36.5, (c) 51.8.

1.16 The mean change is $0.5 \times 10 = 5$. The standard deviation of the change is $2\sqrt{10} = 6.32$.

1.17 VaR is 35. This is because the probability that the loss will fall between 35 and 40 is 5/100 = 5%. Expected shortfall is 37.5. Conditional on the loss being greater than 35, the expected loss is halfway between 35 and 40, or 37.5.

1.18 VaR is USD 3 million. Expected shortfall (USD) is $10 \times 0.6 + 3 \times 0.4 = 7.2$.

1.19 Losses (USD) of 20, 13, 9, 6, 2, and −2 have probabilities of 0.0009, 0.0042, 0.054, 0.0049, 0.126, and 0.81, respectively. The VaR is 9 and ES is

$$\frac{0.0009 \times 20 + 0.0042 \times 13 + (0.05 - 0.0009 - 0.0042) \times 9}{0.05}$$

$= 9.534$

1.20 VaR does not satisfy the subadditivity condition because the VaR for two portfolios combined (9) is greater than the sum of the VaR for each portfolio individually (i.e., $9 > 3 + 3$). Meanwhile, expected shortfall does satisfy the condition because its value for the two portfolios combined is less than the sum of each portfolio's expected shortfall (i.e., $9.534 < 7.2 + 7.2$).

2

Calculating and Applying VaR

Learning Objectives

After completing this reading, you should be able to:

- Explain and provide examples of linear and non-linear portfolios.
- Describe and explain the historical simulation approach for calculating VaR and ES.
- Describe the delta-normal approach and calculate VaR for non-linear derivatives using delta-normal approach.
- Describe and calculate VaR for linear derivatives.
- Describe the limitations of the delta-normal method.
- Explain the Monte Carlo simulation method for calculating VaR and ES and identify its strengths and weaknesses.
- Describe the implications of correlation breakdown for a VaR or ES analysis.
- Describe worst-case scenario analysis and compare it to VaR.

In Chapter 1, we introduced two risk measures: value-at-risk (VaR) and expected shortfall (ES). This chapter discusses how they can be calculated. One popular approach is a non-parametric method, where the future behavior of the underlying market variables is determined in a very direct way from their past behavior. This method is known as *historical simulation* and it enables the development of many future scenarios. These scenarios can be evaluated using full revaluation (which can be computationally time consuming) or the delta-gamma approach.

Models can also be used to calculate risk measures. For a portfolio that is linearly dependent on underlying market variables, it is relatively easy to calculate the mean and standard deviation of changes in the portfolio value using the mean and standard deviation of the changes in those variables. If we assume the returns on the underlying variables (or changes in their values) are multivariate normal, then changes in portfolio value are also normally distributed and thus calculating VaR is relatively straightforward. This approach is sometimes referred to as the *delta-normal model*. For a portfolio that is not linearly dependent on the underlying market variables (e.g., because it contains options), the delta-normal model can also be used. However, it is then less accurate.

An alternative to historical simulation and the delta-normal model is provided by Monte Carlo simulation. This is where scenarios are generated randomly rather than being determined directly from the behavior of market variables in the past.

2.1 LINEAR VERSUS NONLINEAR PORTFOLIOS

A linear portfolio is linearly dependent on the changes in the values of its underlying variables. A simple example of a linear portfolio is one that consists of 100 shares worth USD 50 each. The change in the value of the portfolio (ΔP) is linearly dependent on the change in the stock price (ΔS):

$$\Delta P = 100\Delta S$$

The value of the portfolio is USD 5,000. If Δr is the return on the stock, the previous equation becomes

$$\Delta P = 5000\Delta r$$

This shows there is also a linear relationship between ΔP and Δr.

Now consider a portfolio consisting of long and short positions in stocks. Suppose n_i is the number of shares of stock i in the portfolio (for a short position n_i is negative). If S_i is the price of stock i, the amount invested in stock i is n_iS_i and the increase in the value of the portfolio when the stock price changes by ΔS_i is

$$\Delta P = \sum_i n_i \Delta S_i$$

This can also be written as follows:

$$\Delta P = \sum_{i=1}^{n} q_i \Delta r_i$$

where q_i ($= n_iS_i$) is the amount invested in stock i and Δr_i ($= \Delta S_i/S_i$) is the return on stock i. Again, we see that the change in the portfolio value is a linear function of either the change in stock price (ΔS_i) or the returns (Δr_i).

Portfolios containing more complicated securities are not necessarily linear. Consider, for example, a portfolio consisting of a call option on a stock. The payoff from the option is non-linear (as indicated in Figure 2.1). If the final stock price is less than the strike price, the payoff is zero. When the stock price S is above the strike price X, the payoff is $S - X$.

Prior to maturity, the option price is also a non-linear function of the stock price (as indicated in Figure 2.2). This means that there is *not* a linear relationship between the change in the portfolio value and the change in the stock price.

Assume that the stock price is currently S. The gradient shown in Figure 2.2 gives a good approximation for the relationship between very small changes in the stock price and the change in the option value. But as the stock price changes become larger, the approximation becomes less accurate.

Not all derivatives have the non-linear property illustrated in Figure 2.2. A forward contract on an asset has a value that is a linear function of the asset. If the asset provides no income, the value of a forward contract where the holder is obligated to buy the asset at a future time T for price K is

$$S - \text{PV}(K)$$

where S is the current asset price and PV denotes the present value calculated from the future time T to today. It is easy to

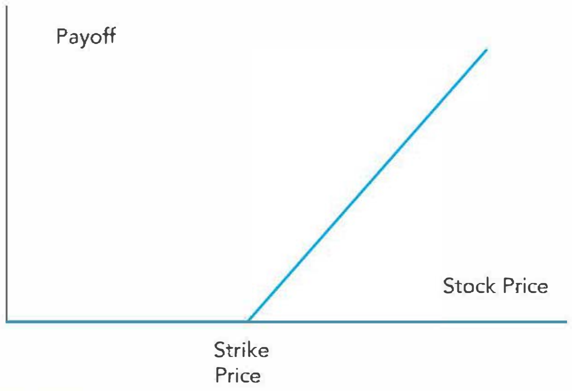


Figure 2.1 **Payoff from a call option at maturity.**

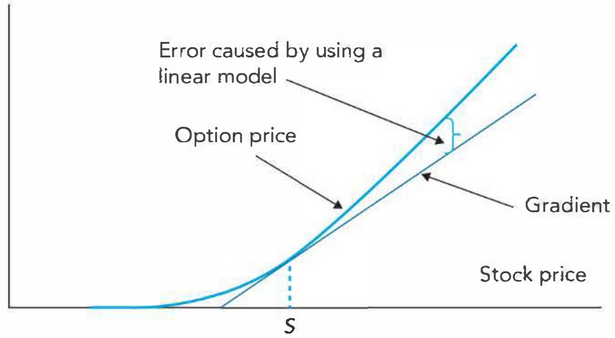


Figure 2.2 Non-linear value of an option prior to maturity. Current stock price is *S*. Assuming linearity introduces some error in the estimation of changes in the value.

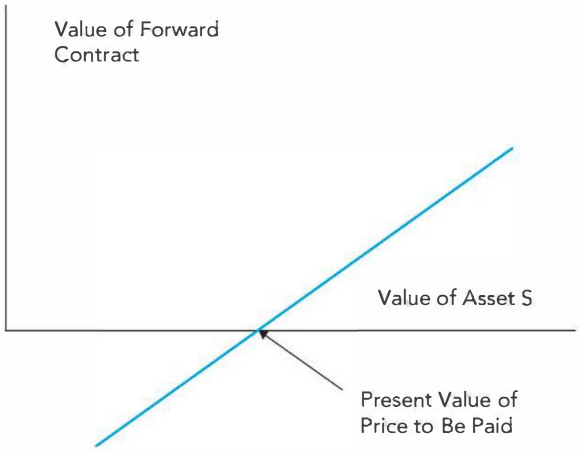


Figure 2.3 Value of forward contract as a function of value of the underlying asset.

see why this formula is correct. If someone has agreed to buy an asset at a future time, this person is essentially in the position of owning the asset today but paying for it at a future time. The value of owning the asset today is S. The present value of what will be paid for the asset is PV(K). Figure 2.3 shows the value of the forward contract as a function of the asset price S.

2.2 HISTORICAL SIMULATION

Historical simulation is a popular method of calculating VaR and ES. Typically, the time horizon is chosen as one day. For longer time horizons, popular assumptions are

$$\text{VaR}(T, X) = \sqrt{T} \times \text{VaR}(1, X)$$

$$\text{ES}(T, X) = \sqrt{T} \times \text{ES}(1, X)$$

where VaR(T, X) and ES(T, X) are the value at risk and expected shortfall (respectively) for a time horizon of T days and confidence level of X. These estimates assume the changes in portfolio value on successive days are normally distributed with a mean of zero and that the changes are independent of each other. In many circumstances, these are reasonable approximations.

Historical simulation involves identifying market variables on which the value of the portfolio under consideration depends. These market variables typically include equity prices, commodity prices, interest rates, credit spreads, volatilities, and so on. The market variables are usually termed *risk factors*.

Daily data is collected on the behavior of the risk factors over a period in the past. (We assume in this section that the past period is one from the immediate past. Later we will explain how other periods are sometimes chosen.) Scenarios are then created by assuming that the change in each risk factor over the next day corresponds to a change observed during one of the previous days.

Table 2.1 shows some data to illustrate the historical simulation procedure. The portfolio under consideration is assumed to depend on many risk factors. The first four are a stock price, an exchange rate, an interest rate, and a credit spread. The interest rate is expressed in percentage terms, whereas the credit spread is expressed in basis points (where one basis point is 0.01%). We assume that today is Day 500, and we are interested in what might happen between today and tomorrow (i.e., between Day 500 and Day 501). The table shows an extract from the most recent 501 days (labeled Day 0 to Day 500) of historical data. This data is used to create 500 scenarios. (We need 501 days of data to produce 500 scenarios because we are looking at the changes from one day to the next.)

In practice, risk factors are divided into two categories:

1. Those where the percentage change in the past is used to define a percentage change in the future, and
2. Those where the actual change in the past is used to define an actual change in the future.

Stock prices and exchange rates are examples of risk factors usually considered to be in the first category. Interest rates and credit spreads are examples of risk factors usually considered to be in the second category.

In the first scenario, we assume all the risk factors behave between Day 500 and Day 501 in the same way as they did between Day 0 and Day 1. As shown in Table 2.1, the stock price increased by 4% between Day 0 and Day 1. For the first scenario that is constructed, we therefore assume that it also increases by 4% between Day 500 and Day 501 so that the Day 501 stock price is USD 65.52 (= 63 × 1.04). Similarly, the exchange rate increased by 0.1% between Day 0 and Day 1. It is therefore

Table 2.1 Historical Data for Example

Day	Stock Price (USD)	Exch. Rate USD/EUR	Int. Rate (%)	Cred. Spr. (bp)	..		..	Port. Val. (USD mill.)
0	50	1.2000	2.52	50	..		..	72.1
1	52	1.2012	2.54	53	..		..	72.5
2	46	1.2015	2.55	54	..		..	70.4
..	..	..	..	..	..		..	
..	..	..	..	..	..		..	
498	60	1.2520	2.30	48				
499	60	1.2550	2.32	46	..		..	75.3
500	63	1.2500	2.36	47	..		..	76.3

Table 2.2 Day 501 Scenarios Created for the Example

Scenario	Stock Price (USD)	Exch. Rate	Int. Rate (%)	Cred. Spr. (bp)	..	..	Port. Val. (USD mill.)	Loss (USD mill.)
1	65.52	1.2513	2.38	50	..	..	76.8	−0.5
2	55.73	1.2503	2.37	48	..	..	71.7	4.6
..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..
499	63.00	1.2530	2.38	45	..	..	75.3	1.0
500	66.15	1.2450	2.40	48	..	..	76.7	−0.4

assumed to also increase by 0.1% between Day 500 and Day 501 so that the Day 501 exchange rate is 1.2513 ($= 1.2500 \times 1.001$).

Meanwhile, the interest rate increased from 2.52% to 2.54% between Day 0 and Day 1. (We consider the actual change of 0.02% rather than the proportional change of 0.02/2.52 for interest rates, as mentioned earlier.) Thus, for the first scenario, we assume the same increase between Day 500 and Day 501 so that the Day 501 interest rate is 2.38% ($= 2.36\% + 0.02\%$). Similarly, the credit spread increased by 3 basis points between Day 0 and Day 1. (Again we consider actual changes rather than proportional changes.) The credit spread is therefore assumed to increase from 47 basis points on Day 500 to 50 ($= 47 + 3$) basis points on Day 501.

The values of the risk factors for the second scenario are calculated similarly using the changes that took place between Day 1 and Day 2. For example, the stock price is USD 55.73 ($= (63 \times (46/52)$). The exchange rate is 1.2503 ($= (1.2500 \times (1.2015/1.2012)$). The interest rate is 2.37%($= 2.36\% + 0.01\%$). The credit spread is 48 ($= 47 + 1$) basis points. The third scenario is similarly calculated from the changes between Day 2 and Day 3, and so on.

Table 2.2 shows the output that is produced at the next stage of the historical simulation. The stock price, exchange rate, interest rate, and credit spread columns are calculated from Table 2.1 as described. We suppose that the current value of the portfolio is USD 76.3 million. To complete the last two columns of the table, it is necessary to produce portfolio values for each of the scenarios and then to calculate losses. (Gains are recorded as negative losses to create a loss distribution.) We assume that the value of the portfolio for Scenario 1 is USD 76.8 million and so there is a gain of USD 0.5 ($= 76.8 - 76.3$) million (which we record as a loss of minus USD 0.5 million). We also assume that the value of the portfolio for the second scenario is 71.7 and therefore a loss of USD 4.6 ($= 76.3 - 71.7$) million.

It is then necessary to sort the losses from the largest to the smallest as presented in Table 2.3.

Suppose that we are interested in the VaR and expected shortfall with a one-day time horizon and a 99% confidence level. The

Table 2.3 Scenarios Sorted from the Worst to the Best

Scenario Number	Loss (USD millions)
210	7.8
195	6.5
2	4.6
23	4.3
48	3.9
367	3.7
235	3.5
..	..
..	..
..	..

VaR can be calculated as the fifth worst loss because the fifth worst loss defines the loss that is at the first percentile point of the distribution ($5/500 = 0.01$).[1] When the VaR percentage applied to the number of observations is not an integer, we can instead interpolate. For example, if there are 250 observations, the 99% VaR level can be assumed to be halfway between the second- and the third-worst loss.

The expected shortfall is usually calculated as the average of the losses that are worse than the VaR level.[2] In our example, the VaR is USD 3.9 million and the expected shortfall is

$$\frac{7.8 + 6.5 + 4.6 + 4.3}{4} = 5.8$$

or USD 5.8 million.

2.3 PORTFOLIO VALUATION

A full revaluation of the portfolio for each of the 500 scenarios can be computationally time consuming. One way of speeding up these calculations is to use what are termed Greek letters. These are hedge parameters used by traders to quantify and manage risks (they will be explained in more detail in a later chapter). The most important Greek letter is delta and it is defined as:

$$\delta = \frac{\Delta P}{\Delta S}$$

where ΔS is a small change in the risk factor and ΔP is the resultant change in the portfolio. Thus, delta is the rate of change in the value of the portfolio with respect to rate of change in the risk factor.

Suppose the risk factor corresponds to a stock price and the delta of the portfolio with respect to that stock price is USD 50,000. This means that the value of the portfolio increases (decreases) by USD 50,000 for each USD 1 increase (decrease) in the stock price. If the stock price increased by USD 3.2, for example, we would estimate a USD 160,000 increase in the portfolio value. In general, the change in the portfolio value arising from the change ΔS in a risk factor is

$$\Delta P = \delta \Delta S \tag{2.1}$$

When there are many risk factors (as in our example), it is necessary to calculate the effect of the change in each one and add them together. Suppose that δ_i is the delta with respect to the ith risk factor and ΔS_i is the change in the ith risk factor for the scenario being considered. The change in the portfolio value is then:

$$\Delta P = \sum_i \delta_i \Delta S_i \tag{2.2}$$

To illustrate this, consider the second scenario in Table 2.2 and suppose the stock price, exchange rate, interest rate, and credit spread are the only risk factors. Suppose further that the deltas calculated for the risk factors (respectively) are 0.5, 2000, -10, and 0.1 (measured as millions of USD per unit of the risk factor). Scenario 2 involves a change in the stock price of -7.27 ($= 55.73 - 63$), a change in the exchange rate of $+0.0003$ ($= 1.2503 - 1.2500$), a change in the interest rate of $+0.01$ ($= 2.37 - 2.36$), and a change in the credit spread of $+1$ ($= 48 - 47$). The estimated change in the portfolio value is therefore

$$(-7.27) \times 0.5 + 0.0003 \times 2{,}000 + 0.01 \times (-10) + 1 \times 0.1 = -3.035$$

or a loss of about USD 3.0 million. For a linear portfolio, delta provides an exact answer. For non-linear portfolios, the change in the portfolio value given by delta is approximate. (The discrepancy between the USD 4.6 million loss in Table 2.2 and USD 3.035 million loss here might indicate that the portfolio in Table 2.2 is non-linear or that there are more than four risk factors in the portfolio.)

The calculation for a non-linear portfolio can be made more accurate by using another Greek letter that measures the curvature in the relationship between the portfolio value and risk

[1] This is the approach we will use and is the one most commonly used, but it is worth noting that there are alternatives. A case could be made for setting VaR in this example equal to the sixth worst loss or an average of the fifth and sixth worst losses. In Excel's PERCENTILE function, when there are n observations and k is an integer, the $k/(n - 1)$ percentile is the observation ranked $k + 1$ and other percentiles are calculated using interpolation.

[2] This is the approach we will use. An alternative is to set the expected shortfall equal to the average of losses that are *less than or equal* to the VaR level. In our example, expected shortfall would then be set equal to the average of the five worst losses.

factors.[3] This letter is called gamma and a better approximation than Equation (2.1) is:

$$\Delta P = \delta \Delta S + \frac{1}{2}\gamma(\Delta S)^2 \qquad (2.3)$$

where γ is the gamma of the risk factor. When there are several risk factors (and each part of the portfolio depends on only one of them) the approximation in Equation (2.3) becomes

$$\Delta P = \sum_i \delta_i \Delta S_i + \tfrac{1}{2}\sum_i \gamma_i(\Delta S_i)^2$$

A more complicated expression involving cross gammas is necessary for situations where there are portfolio components that are dependent on several risk factors.

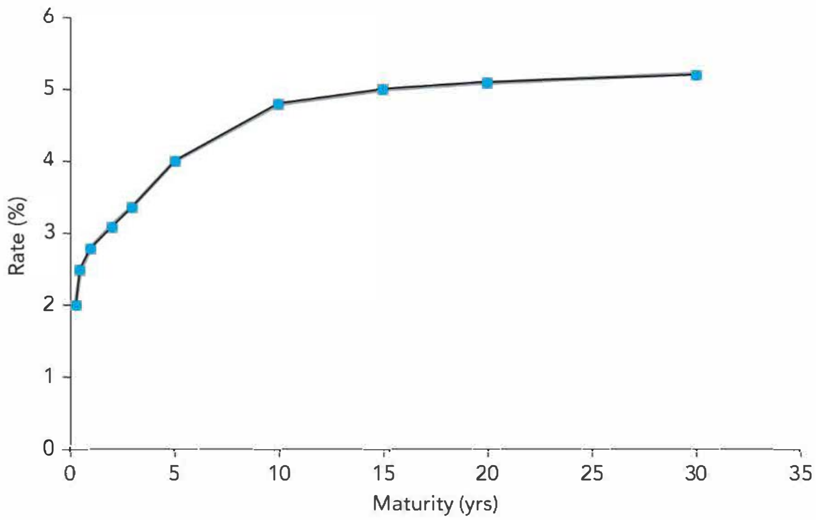


Figure 2.4 An interest rate term structure defined by 10 points.

2.4 TERM STRUCTURES

Risk factors such as stock prices or commodity prices can be described using a single number and are relatively easy to monitor. Other risk factors require a term structure.

Consider, for example, an interest rate such as the U.S. Treasury rate (i.e., the interest rate at which the U.S. government borrows money). This is usually described by a set of points: the one-month rate, three-month rate, six-month rate, and so on. Linear interpolation is commonly used to determine interest rates between these points as illustrated in Table 2.4.

Suppose that the six-month rate is 2.5% and the one-year rate is 2.8%. The nine-month rate would be assumed to be 2.65%, the seven-month rate would be assumed to be 2.55%, and so on. If a historical simulation assumes the six-month rate increases by 10 basis points and the one-year rate increases by 8 basis points, the nine-month rate would be assumed to increase by 9 basis points. (Note that the term structure is usually assumed to be horizontal up to the shortest maturity considered and beyond the longest maturity considered.)

Credit spreads are the excess borrowing rate of a company over a benchmark (such as the Treasury rate). They are defined by a term structure and handled in a similar way to interest rates. The credit spread corresponding to borrowing by a company might be 100 basis points for a one-year maturity, 150 basis points for a five-year maturity, and 180 basis points for a ten-year maturity. This means that the company pays 1% more than the risk-free rate when it borrows funds for one year, 1.5% more than the risk-free rate when it borrows funds for five years, and 1.8% more than the risk-free rate when it borrows funds for ten years. Linear interpolation would be used to determine intermediate points. For example, the credit spread for seven years would be assumed to be 162 basis points.

A further risk factor that is defined by a term structure is an at-the-money implied volatility. This is the volatility that must be substituted into the Black-Scholes Merton formula to get the market price of an option. (This concept will be discussed in a later chapter).

2.5 STRESSED MEASURES

In the previous example, the data used to develop scenarios was from an immediately preceding period. However, bank regulators have moved to using what are termed stressed VaR and stressed expected shortfall for determining bank capital requirements. These are measures based on data from a one-year period that would be particularly stressful for a bank's current portfolio. For example, March 2008 to February 2009 could be chosen by a bank that would be badly impacted by a crisis like the one experienced in 2007–2008.

The scenarios are calculated from the changes that took place during the stressed period. Typically, 250 scenarios (produced from one year of data) are used rather than 500 scenarios (produced from two years of data). Everything else about the calculations remains the same. For example, if the percentage change in an exchange rate during a day in the stressed

[3] The gamma for a risk factor can be defined as the rate of change of delta with respect to the risk factor. Readers familiar with calculus will recognize Equation (2.3) as an approximation from a Taylor Series expansion:

$$\Delta P = \delta\frac{\partial P}{\partial S} + 0.5\gamma\frac{\partial^2 P}{\partial S^2}$$

period was 5%, the scenario corresponding to that day would assume a 5% change in the exchange rate between today and tomorrow.

2.6 THE DELTA-NORMAL MODEL

An alternative to historical simulation is the use of the delta-normal model. This is based on Equation (2.2):

$$\Delta P = \sum_i \delta_i \Delta S_i$$

As explained, this is an exact result in the case of a linear portfolio and an approximate result for non-linear portfolios.

As discussed earlier, there are two categories of risk factors:

1. Those such as stock prices and commodity prices where it is natural to consider percentage changes in the risk factor, and
2. Those such as interest rates and credit spreads where it is natural to consider actual changes in the risk factor.

To accommodate both types of risk factors it is convenient to rewrite Equation (2.2) as

$$\Delta P = \sum_i a_i x_i \quad (2.4)$$

In the case of risk factors where percentage changes are considered, $x_i = \Delta S_i / S_i$ and $a_i = \delta_i S_i$. In the case of risk factors where actual changes are considered, $x_i = \Delta S_i$ and $a_i = \delta_i$.[4]

The mean (μ_P) and standard deviation (σ_P) of ΔP can be calculated from Equation (2.4):

$$\mu_P = \sum_{i=1}^{n} a_i \mu_i$$

$$\sigma_P^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \rho_{ij} \sigma_i \sigma_j$$

where μ_i and σ_i are the mean and standard deviation of x_i (respectively) and ρ_{ij} is the coefficient of correlation between x_i and x_j. The expression for σ_P can be written as

$$\sigma_P^2 = \sum_i a_i^2 \sigma_i^2 + 2\sum_{i>j} a_i a_j \rho_{ij} \sigma_i \sigma_j$$

At this stage, we have made no assumptions about the probability distributions of the risk factor changes. All we have done is show that if we estimate the means of, standard deviations of, and correlations between the risk factors, we can calculate the mean and standard deviation of the change in the value of a linear portfolio.

[4] For risk factors in the first category, practitioners often choose to redefine delta as the sensitivity of the value of the portfolio to a small percentage change in the risk factor. The a_i's are then all deltas.

If we now assume that the change in the portfolio value is normal, it is easy to calculate either VaR or expected shortfall. As discussed in Chapter 1, VaR is

$$\mu_P + \sigma_P U$$

Meanwhile, expected shortfall is

$$\mu_P + \sigma_P \frac{e^{-U^2/2}}{(1 - X)\sqrt{2\pi}}$$

where X is the confidence level and U is the point on the normal distribution that has an X probability of being exceeded. For example, if the confidence level X is 95%, then $U = -1.645$. If X is 99%, then $U = -2.326$.

The change in the value of a linear portfolio is normally distributed if each of the x_i is normal. This means that for a risk factor such as an equity price, the return over the period considered is normal. For a risk factor such as an interest rate, it means that the changes are normal.

A portfolio may also be approximately normal in other situations. This is because when many non-normal variables are added, a theorem in statistics known as the Central Limit Theorem (CLT) tells us that there is a tendency for the result to be approximately normal. Specifically, the CLT applies when the variables are independent and random (conditions which may or may not hold in a given situation).

A common assumption is that the mean change in each risk factor is zero and therefore the average change in a linear portfolio is also zero. The assumption is not exactly true, but it is reasonable when considering short time periods (because the mean change in the value of a portfolio is much less than its standard deviation during a short time period). This assumption simplifies the expression for VaR to:

$$\sigma_P U$$

Meanwhile, the expression for expected shortfall simplifies to:

$$\sigma_P \frac{e^{-U^2/2}}{(1 - X)\sqrt{2\pi}}$$

The model we have been discussing is referred to as the delta-normal model because it is based on the risk factor deltas of a portfolio and assumes normality.

2.7 LIMITATIONS OF DELTA-NORMAL MODEL

The delta-normal model works well for linear portfolios when the risk factor probability distributions are at least approximately normal. It can also be used for non-linear portfolios. However, Equation (2.4) is then an approximation. We illustrated this with

Table 2.4 Accuracy of the Delta-Normal Model for Option A (An at-the-Money Option) and Option B (An in-the-Money Option) When the Stock Price Increases from 100 to 105

Option	Delta	Gamma	Initial Value	Change in Value Predicted by Delta	Actual Change in Value
A	0.5299	0.0265	5.9785	2.6495	2.9701
B	0.9409	0.0078	20.4036	4.7046	4.7860

a call option in Figure 2.2. To understand why this is so, we return to Equation (2.3):

$$\Delta P = \delta \Delta S + \frac{1}{2}\gamma(\Delta S)^2$$

The parameter γ is a measure of curvature (or non-linearity). If γ is small, the linear model is a reasonable approximation. But as γ becomes bigger, this is no longer true.

We can illustrate this by considering two call options on a stock worth USD 100. Option A is at-the-money (i.e., it has a strike price of USD 100) while Option B is in-the-money with a strike price of USD 80. Both options last for three months. We assume that the volatility of the stock price is 30% and the interest rate is zero. In Table 2.4, we suppose the stock price increases from USD 100 to USD 105. The last two columns in Table 2.4 show the estimated change in the options prices (as predicted by delta) and actual change in the options prices.

Table 2.4 shows that delta does not give a good estimate for the price change of the at-the-money option. However, it works much better for the price change of the in-the-money option. This is not surprising because the gamma for the at-the-money option is much higher than that of the in-the-money option.

When the gamma effect is considered using Equation (2.3), price change predictions improve considerably. For Option A, the prediction is

$$0.5299 \times 5 + 0.5 \times 0.0265 \times 25 = 2.9810$$

For Option B, the prediction is

$$0.9409 \times 5 + 0.5 \times 0.0078 \times 25 = 4.8027$$

As we saw earlier, not all derivatives are non-linear. A forward contract, for example, is a linear product. (One reason why the delta-normal model works so well for the in-the-money option in Table 2.4 is that the option is almost certain to be exercised and is therefore similar to a forward contract to buy the stock for USD 100.)

Sometimes the impact of a change in the underlying risk factor on the value of the portfolio is highly non-linear. An example here is provided by a butterfly spread. This is a position in three options that produces a modest positive payoff if there is a small change in the risk factor and a small loss if there is a large change.

In the case of a linear product, a normal distribution assumption for the change in a risk factor translates into a normal distribution for the change in the portfolio value. This is not the case for a non-linear product. Figure 2.5 shows how a normal distribution for the asset price (shown next to the horizontal axis) translates into a skewed distribution for the call option (shown next to the vertical axis).

The gradient in Figure 2.6 uses the delta-normal model to translate the normal distribution for the asset price into a normal distribution for the option price. Comparing Figures 2.5 and 2.6, we see that the delta-normal model understates the probability of high option values and overstates the probability of low option values. For a long position in the call option, the delta-normal model will therefore give an excessively high value for the VaR and the expected shortfall. For a short position in the call option, the VaR and the expected shortfall will be too low.

In practice, the discrepancies described in the previous paragraphs can create serious problems for risk managers. Many derivatives portfolios are kept close to delta neutral as a matter of course. (As we will explain later in this book, traders often must take positions in the underlying risk factors at the end of a day to bring their deltas close to zero.) This means that the

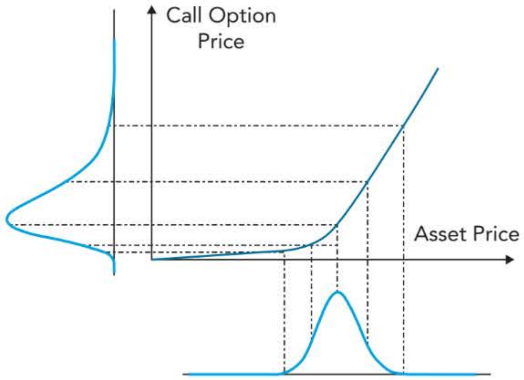


Figure 2.5 Illustration of the fact that a normal distribution for the asset price translates into a skewed distribution for the value of the call option.

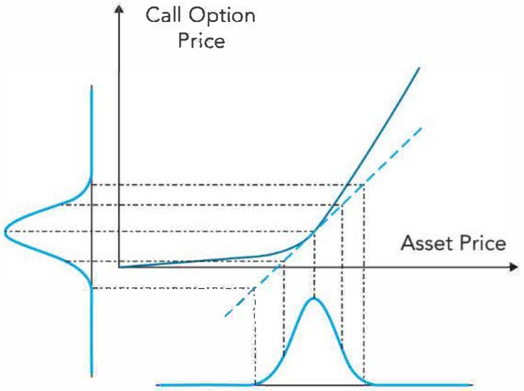


Figure 2.6 **Translation of price of asset to price of option for the delta-normal model.**

non-linear (curvature) risk is the most important residual risk. And as we have seen, this risk is not considered by the delta-normal model.

As we saw earlier, using delta and gamma to predict changes can work much better than just using delta. The resulting model is a quadratic model approximation (rather than a linear model approximation). Unfortunately, while there are easy-to-use analytic results for the delta approximation, there are no similar results for the quadratic approximation.

2.8 MONTE CARLO SIMULATION

Another way to calculate VaR and expected shortfall is to use Monte Carlo simulation. This approach is very similar to the use of historical simulation, but with one key difference: Monte Carlo simulation generates scenarios by taking random samples from the distributions assumed for the risk factors (rather than using historical data).[5] Monte Carlo simulation works for both linear and non-linear portfolios.

Suppose that we assume the risk factor changes have a multivariate normal distribution (as in the delta-normal model). The procedure is as follows:

1. Value the portfolio today using the current values of the risk factors
2. Sample once from the multivariate normal probability distribution for the Δx_i. The sampling should be done in a way consistent with the assumed standard deviations and correlations (usually estimated from historical data). As we indicated earlier, for risk factors such as equity prices and exchange rates, the Δx_i are percentage changes; for risk factors such as interest rates and credit spreads, they are actual changes.
3. Use the sampled values of the Δx_i to determine the values of the risk factors at the end of the period under consideration (usually one day)
4. Revalue the portfolio using these risk factor values
5. Subtract this portfolio value from the current value to determine the loss
6. Repeat steps 2 to 5 many times to determine a probability distribution for the loss

Suppose that there is a total of 1000 Monte Carlo simulation trials. The VaR with a 99% confidence level for the period considered will be the tenth worst loss and the expected shortfall is the average of the nine losses worse than this. As with other methods for calculating VaR and expected shortfall, the period considered when the simulation is carried out is usually one day and the equations given earlier:

$$\text{VaR}(T, X) = \sqrt{T} \times \text{VaR}(1, X)$$
$$\text{ES}(T, X) = \sqrt{T} \times \text{ES}(1, X)$$

are used to convert the measure with a one-day time horizon to the corresponding measure with a T-day time horizon.

Unfortunately, Monte Carlo simulation is computationally intensive and thus quite slow. The portfolios that must be assessed are often quite large and a full revaluation of them on each simulation trial can be time consuming. As in the case of historical simulation, the computational time can be reduced by using the delta-gamma approach in Section 2.3. This is sometimes referred to as using *partial simulation*.

In the delta-normal approach it is necessary to assume normal distributions for the risk factors. When Monte Carlo simulation is used, any distribution can be assumed for the risk factors providing correlations between the risk factors can be defined in some way. One approach can be to sample from a multivariate normal distribution in a way that reflects historical correlations and then transform the sampled values to the non-normal distributions that are considered appropriate. The transformation is accomplished on a "percentile-to-percentile" basis. If the value sampled from the normal distribution is the p-percentile, it is transformed to the p-percentile of the assumed distribution.[6]

[5] Assuming distributions for the risk factors is sometimes referred to as a using a structured Monte Carlo simulation.

[6] See J. Hull and A. White, "Value at risk when daily changes are not normally distributed," *Journal of Derivatives*, 5, 3, (Spring 1998): 9–19. This is a Gaussian copula approach to defined correlations between non-normal distributions and will be discussed further in chapter 6.

2.9 ESTIMATING PARAMETER VALUES

To implement the delta-normal approach or Monte Carlo simulation, it is necessary to estimate the standard deviations and correlations for either proportional or actual changes in the risk factors. This will typically be done using historical data. The most natural approach is to use data from a recent period. The exponentially weighted moving average and GARCH models, which will be discussed in the next chapter, can be used to apply more weight to recent observations than to observations from long ago.

As discussed in Section 2.5, bank regulators like to use "stressed VaR" and "stressed expected shortfall" when determining bank regulatory capital. To determine these measures, it is necessary to estimate standard deviations and correlations from a period in the past that would be particularly stressful for the current portfolio.

2.10 CORRELATION BREAKDOWN

Not surprisingly, standard deviations increase during stressed market conditions. Interestingly, it is also true that correlations generally increase. It is sometimes stated that "in stressed markets all correlations go to one." This is, of course, an exaggeration, but the phenomenon that correlations increase in stressed markets appears to be real. For example, during the 2007–2008 crisis, default rates on mortgages in all parts of the United States increased together. When there is a flight to quality in bond markets, the credit spreads of all bonds throughout the world tend to rise together.

The key point here is that, in periods of heightened volatility, correlations can be quite different from those in normal market conditions. This is sometimes referred to as a *correlation breakdown*. It does have significant implications for risk managers. When calculating VaR or ES they are concerned with estimating what will happen in extreme market conditions. They should, therefore, try and estimate what correlations will be in such conditions rather than what they are in normal market conditions.

2.11 WORST CASE ANALYSIS

Occasionally, when there are repeated trials, an analyst will calculate statistics for worst-case results. For example, if a portfolio manager reports results every week, he or she might ask what the worst result will be over a period of 52 weeks. If the distribution of the returns in one week is known, a Monte Carlo simulation can be used to calculate statistics for this worst-case result. For example, one can calculate the expected worst-case result over 52 weeks, the 95th percentile of the worst-case result, and so on.

However, these worst-case statistics may be overly pessimistic and should not usually be regarded as alternatives to VaR and ES.

SUMMARY

There are several different ways of calculating risk measures such as VaR and expected shortfall. Historical simulation is one of the most popular methods. It assumes that what will happen over the next day is a random sample of what has happened over previous days. The advantage of this approach is that it automatically reflects the means of, standard deviations of, and correlations between risk factors that have been observed in the past.

For linear portfolios, the delta-normal model is a popular approach. It produces answers very quickly because calculations are carried out using an analytic formula. The approach is closely related to the pioneering work of Markowitz and often used by portfolio managers when their portfolios consist predominantly of long and short positions in stocks and other assets. Unfortunately, the approach does not work well for a portfolio containing options. This is particularly true if the portfolio is delta-neutral, as is often the case for a derivatives trader.

Monte Carlo simulation is another approach. It is like historical simulation, except that scenarios are generated by sampling from estimates of the distributions for risk factors. They are not based directly on historical movements.

Since the 2008–2009 crisis, bank regulators have moved toward using stressed VaR and stressed expected shortfall. These are measures calculated from the way risk factors behaved during a stressed period rather than during the immediately preceding period.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

QUESTIONS

Short Concept Questions

2.1 What is the difference between a linear and a non-linear portfolio?

2.2 What is the relationship usually assumed between the VaR with 99% confidence for a ten-day time horizon and the VaR with 99% confidence for a one-day time horizon?

2.3 Do analysts model percentage changes or actual changes for (a) exchange rates, (b) interest rates, (c) credit spreads?

2.4 If there are 400 simulations on the loss (gain) from an investment, how is VaR with a 99% confidence level calculated?

2.5 In the situation in Question 2.4, how is expected shortfall calculated?

2.6 What does the delta-normal model assume?

2.7 Why is the delta-normal imprecise for non-linear portfolios?

2.8 How is stressed VaR defined?

2.9 What does gamma measure?

2.10 Do correlations tend to increase or decrease during stressed market conditions?

Practice Questions

2.11 A portfolio consists of 100 shares worth USD 20 each and 200 shares worth USD 30 each. What is the change in the value of the portfolio (in USD) as a function of the change in share prices (in USD)?

2.12 In the situation in Question 2.11, what is the change in the value of the portfolio (in USD) as a function of the returns (i.e., the percentage change in share prices)?

2.13 Suppose that the portfolio in Table 2.1 depends on a stock index. The value of the stock index on Days 0, 1, 2, 498, 499, and 500 is 1900, 1930, 1890, 2200, 2250, 2275, respectively. What is the value of the stock index in the corresponding Day 501 scenarios 1, 2, 499, and 500?

2.14 Suppose that the portfolio in Table 2.1 depends on another credit spread. The value of the credit spread on Days 0, 1, 2, 498, 499, and 500 is (in basis points) 190, 195, 205, 170, 172, 176, respectively. What is the value of the credit spread in the corresponding Day 501 scenarios 1, 2, 499, and 500?

2.15 How would you estimate the 99.5% VaR from Table 2.3?

2.16 Consider a position consisting of a USD 10,000 investment in asset X and a USD 20,000 investment in asset Y. Assume that the daily volatilities of X and Y are 1% and 2% and that the coefficient of correlation between their returns is 0.3. What is the five-day VaR with a 97% confidence level?

2.17 In the situation in Question 2.16, what is the five-day expected shortfall with a 97% confidence level?

2.18 A portfolio depends on a 3.5-year, risk-free interest rate and no other rates or risk factors. It is calculated that the portfolio will increase in value by USD 500 for each one-basis-point increase in the rate. The rates that are modeled are the three-month, six-month, one-year, two-year, three-year, five-year, ten-year, and thirty-year rates. Applying linear interpolation, what is the delta of the portfolio with respect to each of these rates?

2.19 A non-linear portfolio depends on a stock price. The delta is 30 and the gamma is 5. Estimate the impact of a USD 2 change in the stock price on the value of the portfolio with all else remaining the same.

2.20 Consider a forward contract to purchase an asset currently worth USD 10,000 for USD 12,000 in three years. The interest rate is 3% with annual compounding. How much is the forward contract worth?

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

ANSWERS

Short Concept Questions

2.1 A linear portfolio is linearly dependent on the underlying risk factors whereas a non-linear portfolio is not.

2.2 The ten-day VaR is the square root of 10 multiplied by the one-day VaR.

2.3 (a) percentage, (b) actual, (c) actual

2.4 It would be the fourth worst loss.

2.5 It is the average of the three worst losses.

2.6 The delta normal model assumes a linear relationship between changes in the value of a portfolio and the changes in the value of the portfolio risk factors.

2.7 The delta-normal model does not consider the curvature of the relationship between the portfolio value and individual risk factors.

2.8 Stressed VaR is the VaR calculated using the changes in risk factors that were observed during a period of stressed market conditions.

2.9 Gamma is a measure of the curvature of the relationship between the value of a portfolio and the value of a risk factor.

2.10 Correlations usually increase in stressed market conditions.

Solved Problems

2.11 $\Delta P = 100\Delta S_1 + 200\Delta S_2$ where S_1 and S_2 are the stock prices and P is the portfolio value.

2.12 $\Delta P = 2000\Delta r_1 + 6000\Delta r_2$ where r_1 and r_2 are the returns on the stocks.

2.13 The values of the stock index for scenarios 1, 2 , 499, and 500 are

$$2275 \times 1930/1900 = 2310.9$$

$$2275 \times 1890/1930 = 2227.8$$

$$2275 \times 2250/2200 = 2326.7$$

$$2275 \times 2275/2250 = 2300.3$$

2.14 The values of the stock index for scenarios 1, 2 , 499, and 500 are

$$176 + 195 - 190 = 181$$

$$176 + 205 - 195 = 186$$

$$176 + 172 - 170 = 178$$

$$176 + 176 - 172 = 180$$

2.15 It can reasonably be estimated as half way between the second and third worst loss or USD 5.55 million.

2.16 The standard deviation of the daily changes in the assets are (in USD) 100 and 400. The standard deviation of the daily change in the portfolio is

$$\sqrt{100^2 + 400^2 + 2 \times 100 \times 400 \times 0.3} = 440.5$$

The standard deviation of the five-day change is the square root of 5 multiplied by the one-day standard deviation, which is USD 984.9. The 97% VaR is 1.88 times this, which is USD 1852.4.

2.17 The five-day expected shortfall is

$$984.9\frac{e^{-1.88^2/2}}{0.03\sqrt{2\pi}} = 2233.8$$

2.18 The portfolio has a delta with respect to the 3- and 5-year rates and no other rates. When the 3-year rate changes by one basis point, the 3.5-year rate changes by 0.75 basis points. When the 5-year rate changes by one basis point, the 3.5-year rate changes by 0.25 basis points. The deltas with respect the 3- and 5-year rates are therefore USD 375 and USD 125, respectively.

2.19 The impact (in USD) is $30 \times 2 + 0.5 \times 5 \times 2^2 = 70$.

2.20 The forward contract is worth (in USD) $10{,}000 - (12{,}000/1.03^3) = -981.7$.

3

Measuring and Monitoring Volatility

Learning Objectives

After completing this reading, you should be able to:

- Explain how asset return distributions tend to deviate from the normal distribution.
- Explain reasons for fat tails in a return distribution and describe their implications.
- Differentiate between conditional and unconditional distributions and describe regime switching.
- Compare and contrast different approaches for estimating conditional volatility.
- Apply the exponentially weighted moving average (EWMA) approach to estimate volatility, and describe alternative approaches to weighting historical return data.
- Apply the GARCH (1,1) model to estimate volatility.
- Explain and apply approaches to estimate long horizon volatility/VaR and describe the process of mean reversion according to a GARCH (1,1) model.
- Evaluate implied volatility as a predictor of future volatility and its shortcomings.
- Describe an example of updating correlation estimates.

A constant volatility would be fairly easy to estimate using historical data. In practice, however, volatility changes through time. This leads to situations where asset returns are not normally distributed; instead, they tend to have fatter tails than a normal distribution would predict. This is important for the estimation of risk measures (such as VaR and expected shortfall) because these measures depend critically on the tails of asset return distributions.

An alternative to assuming asset returns are normal with a constant volatility is to assume they are normal conditioned on the volatility being known. When volatility is high, the daily return is normal with a high standard deviation. When the volatility is low, the daily return is normal with a low standard deviation. The conditionally normal model may not be perfect, but it is an improvement over the constant volatility model.

To implement the conditionally normal model (so that it can, for example, be used in conjunction with the delta-normal model introduced in Chapter 2), it is necessary to monitor volatility so that a current volatility estimate is produced. In this chapter we present two ways of doing that: the exponentially weighted moving average (EWMA) model and the GARCH (1,1) model. We also discuss how volatility can be estimated over periods longer than one day and how the tools used to monitor volatility can also be used to monitor correlation.

3.1 DEVIATIONS FROM NORMALITY

There are three ways in which an asset's return can deviate from normality.

1. The return distribution can have fatter tails than a normal distribution.
2. The return distribution can be non-symmetrical.
3. The return distribution can be unstable with parameters that vary through time.

The third situation can lead to either of the first two situations. For example, the return distribution can be unstable when the volatility parameter changes through time. This leads to a return distribution with fatter tails than the normal distribution. It is easy to understand why volatility is non-constant. When markets become stressed, volatility tends to increase. As markets calm, volatility tends to decrease.

To illustrate the fat-tails phenomenon, imagine a simple situation where a variable (X) is a mixture of two normal distributions. During a certain period, X has

- A 50% chance of being normal with a mean of zero and a standard deviation of σ_1, and
- A 50% chance of being normal with a mean of zero and a standard deviation of σ_2.

This is referred to as a *mixture distribution*. The probability density function for a normal distribution was introduced in Chapter 1. The distribution we are considering here has a density function that has a 50% chance of being a normal distribution with mean zero and standard deviation σ_1, and a 50% chance of being a normal distribution with mean zero and standard deviation σ_2. The probability density function is therefore a weighted average of two normal probability density functions:

$$0.5\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-X^2/(2\sigma_1^2)}+0.5\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-X^2/(2\sigma_2^2)}$$

Figure 3.1 plots this mixture distribution and compares it with the density function for a normal distribution (with the same standard deviation) when $\sigma_1 = 0.5$ and $\sigma_2 = 1.5$.[1]

Note the mixture distribution in Figure 3.1 has the same characteristics as the distribution for the returns on the S&P 500 considered in Chapter 1. When it is compared with a normal distribution with the same standard deviation, we find that probability mass has been taken from the area surrounding the one standard deviation point and distributed to the center of the distribution and to the tails. This is what usually happens with a symmetrical fat-tailed distribution. Because the mixture distribution is fat-tailed, there is more probability mass in the tails of the distribution. In order to keep the standard deviation equal to that of the normal distribution, however, we must also have more probability mass in the center of the mixture distribution. In other words, the mixture distribution must be more "peaked" than the normal distribution.

An unstable mean can also lead to a non-normal distribution. The means of distributions can change through time for several reasons. Consider, for example, the expected return on equities. This is the sum of a risk-free rate and a risk premium. In practice, both change through time, and so the expected return also changes through time.

The impact of the mean changing through time can be illustrated in the same way as the impact of the volatility changing through time. Interestingly, combining a stochastic mean with a stochastic volatility can give rise to a non-symmetrical distribution. Consider a mixture of two normal distributions: The first has a mean of 0.3 and a standard deviation of 0.5, whereas the second has a mean of 1.5 and a standard deviation of 2.0. The result is the skewed distribution shown in Figure 3.2.

For most of this chapter, we will focus on the impact of volatility changing through time. For daily data, the volatility is much more important than the mean return. In fact (as we have pointed out in earlier chapters), it is often not too much of an approximation to assume that the mean daily return is zero.

[1] The variance of the normal distribution is the average of 0.5^2 and 1.5^2.

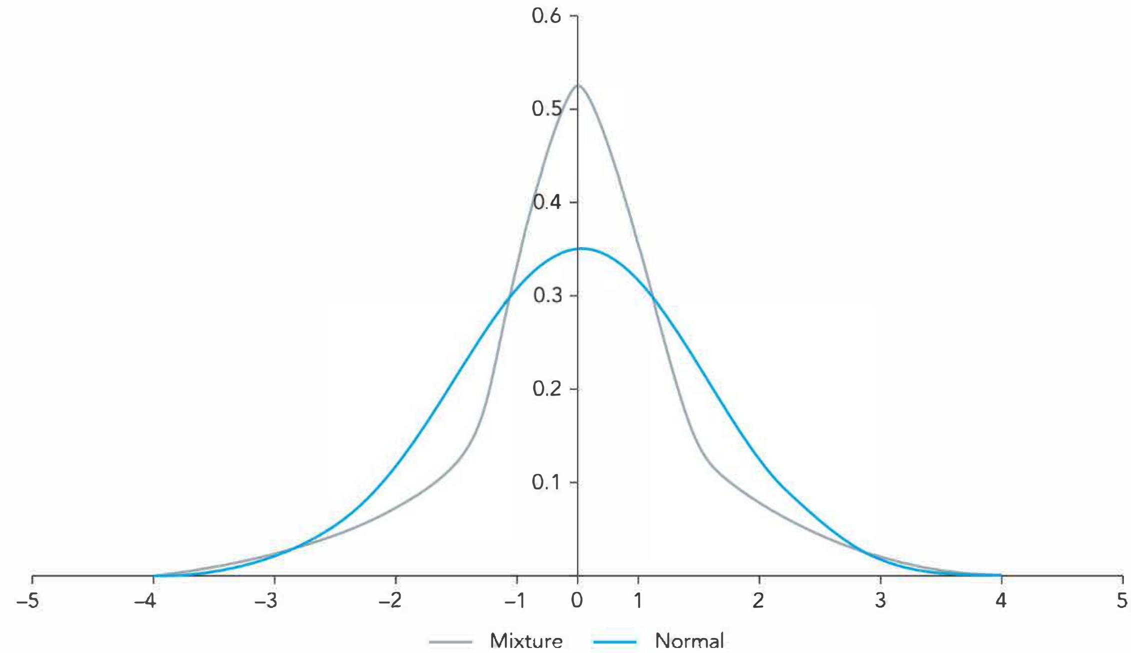


Figure 3.1 A comparison of a mixture of two normal distributions, with a normal distribution that has the same standard deviation.

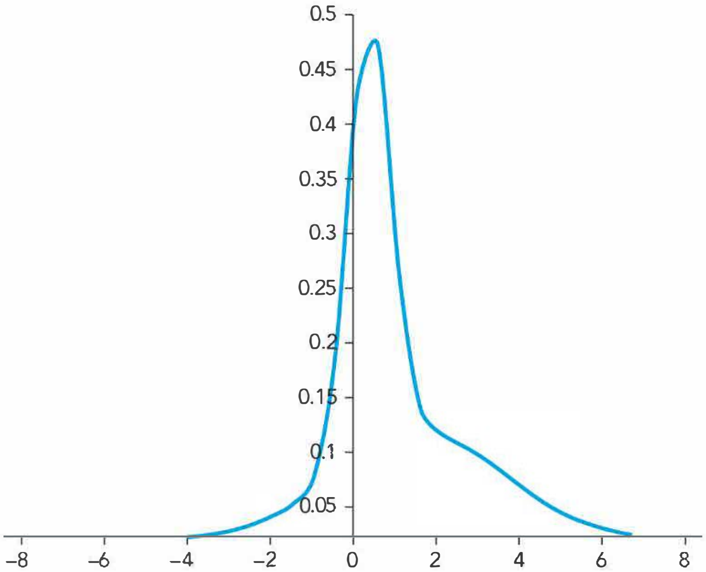


Figure 3.2 Example of non-symmetrical distribution created for a mixture when both means and standard deviations are different.

3.2 UNCONDITIONAL AND CONDITIONAL NORMALITY

To understand how fat tails are created by a stochastic volatility (i.e., a volatility that changes through time in an unpredictable way), it is important to distinguish between a model where returns are unconditionally normal and a model where they are conditionally normal.

In a model where returns are unconditionally normal, the probability distribution of the return each day has the same normal distribution with the same standard deviation.

In a model where the return is conditionally normal, however, the return distribution is normal each day, while the standard deviation of the return varies throughout time. During some periods it is high; during other periods it is low. This leads to an unconditional distribution with fat tails.

A key point here is that when we collect data on daily returns, we observe the unconditional distribution (rather than the conditional distribution). This is why we observe fat tails when the distribution is conditionally normal.

Volatility can be monitored to estimate a conditional distribution for daily returns. Suppose the standard deviation of returns has averaged 1% per day historically, but our procedure for monitoring volatility indicates volatility is currently 2%. Assuming the return distribution is normal with a standard deviation of 2% would yield more accurate measurements for VaR and expected shortfall than using a fat-tailed distribution estimated from historical data and much more accurate results than assuming a normal distribution with a standard deviation of 1%.

Slow Changes versus Regime Switching

It is often reasonable to suppose that volatility changes slowly. To quote famous mathematician Benoit Mandelbrot, "large changes tend to be followed by large changes—of either sign—and small changes tend to be followed by small changes."[2] When volatility is high we see large positive and negative changes in the price of an asset. When volatility is low, the positive and negative changes are much lower. When volatility changes slowly as indicated by Mandelbrot's statement, there are (as we will see) tools that can be used to estimate the current value of volatility (i.e., to monitor volatility).

Sometimes, however, volatility changes abruptly, so that Mandelbrot's statement does not hold. Because of an unexpected event or government action, volatility might jump suddenly from 1% to 3% per day. When markets calm down, it could suddenly jump back to 1% per day. This phenomenon is referred to as *regime switching*. Regime switches present additional challenges for risk managers because they are usually unanticipated. Models that have been working well may suddenly stop working well because the economic environment has changed.

3.3 HOW IS VOLATILITY MEASURED

Up to now we have used the term volatility without defining it precisely. A rough definition is that the volatility of a variable measures the extent to which its value changes through time. For example, a stock price with high volatility exhibits larger daily movements than one with a low volatility.

In risk management, the volatility of an asset is the standard deviation of its return in one day. Denote the return on day i by r_i. Assuming the asset provides no income:

$$r_i = \frac{S_i - S_{i-1}}{S_{i-1}}$$

where S_i is the value of the asset at the close of trading on day i.[3]

The usual formula for calculating standard deviations from sample data would give the volatility estimated for day n (σ_n) from the return on the m previous days as:

$$\sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(r_{n-i} - \bar{r})^2}$$

Here, $\bar{r}$ is the average return over the m previous days:

$$\bar{r} = \frac{1}{m}\sum_{i=1}^{m} r_{n-i}$$

[2] Mandelbrot, B. (1967). The Variation of Some Other Speculative Prices. *The Journal of Business*, *40*(4), 393–413.

[3] This is the usual definition of return used by risk managers. Analysts sometimes calculate the return as $\ln(S_i/S_{i-1})$. This is the return with continuous compounding rather than with daily compounding. When one day periods are considered the two definitions are very similar.

In risk management we usually simplify this formula in two ways:

1. We replace $m - 1$ by m, and
2. We assume that $\bar{r} = 0$.

The first simplification can be justified theoretically. The formula involving $m - 1$ is an unbiased estimate of the volatility. When we replace the $m - 1$ by m, we get a maximum likelihood estimate of the volatility (i.e., an estimate that is most likely given the data).

The second simplification is reasonable.[4] This is because over short periods, the standard deviation of the return is much more important than the mean return. We are also interested in estimating the expected mean return (not the historical mean return). If the historical mean return (i.e., the mean return calculated from the sample data) is high (low), it is not the case that the expected mean return over the next short period of time is high (low).[5]

The two simplifications lead to the formula:

$$\sigma_n^2 = \frac{1}{m}\sum_{i=1}^{m} r_{n-i}^2 \tag{3.1}$$

We define the square of the volatility as the *variance rate*. The formula we have produced shows that the variance rate is the average of squared returns.

Suppose the returns on ten successive days are as shown in Table 3.1. To apply Equation (3.1) we first convert the returns to decimals so that they are 0.011, −0.006, 0.015, ... on days 1, 2, 3 .... We then calculate the sum of squared returns to be 0.001165. Therefore, the average squared return is 0.0001165 (= 0.001165/10). This is the estimate of the variance rate. The volatility per day is 0.0108 (= $\sqrt{0.0001165}$), or 1.08%.

Using the Absolute Value of Returns

It is interesting to mention an alternative procedure for estimating volatility. This involves averaging absolute returns, rather than working with squared returns as in Equation (3.1). This method is considered more robust when the distribution is non-normal. Indeed, averaging absolute returns often provides a superior forecast for fat-tailed distributions. However, the formula in Equation (3.1) is the one most widely used and the one which we will focus on in this chapter.

[4] Usually in risk management we work with estimates of the volatility per day. However, if longer periods are considered, the assumption that the mean is zero may not be appropriate. For example, over one year, the mean return from a stock might be estimated as 12% and the volatility as 20%. The mean should then be taken into account.

[5] A related point here is that a huge amount of historical data is necessary for an accurate estimate of the mean even when the mean remains constant. Standard deviation can be estimated accurately with much less data.

Table 3.1 Data on Daily Returns

Day	Return	Day	Return
1	+1.1%	6	+0.4%
2	−0.6%	7	+1.8%
3	+1.5%	8	−0.3%
4	−2.0%	9	−0.3%
5	+0.3%	10	−0.4%

3.4 ESTIMATING THE CURRENT VOLATILITY

Although not perfect, the conditionally normal model is useful in many circumstances. For a risk manager interested in using the model, it is important to find a way of estimating current volatility. The delta-normal approach (described in Chapter 2) can then be used to estimate VaR or expected shortfall using conditional distributions.

One approach is to estimate volatility using Equation (3.1). However, there are dangers with this method. If m is too large, the estimate of the current volatility will not be reliable because it will use data from too long ago. If $m = 250$, for example, volatility is estimated from daily data collected over one calendar year. If the volatility is cyclical (as it often is), it could be much lower (or higher) during the first six months than it is today. The calculated estimate would then not reflect current volatility. With a smaller value (e.g., $m = 10$ or $m = 20$), the data being used is more relevant to estimating current volatility. However, the resulting volatility estimate may not be very accurate because it only uses a small number of data points.

The standard error of an estimate is the standard deviation of the difference between the estimate and the true value. The standard error of a volatility estimate calculated from m observations is approximately equal to the estimate divided by the square root of $2(m - 1)$. In Table 3.1, $m = 10$ and provides a manageable example of the calculation of volatility. However, the volatility estimate itself has a standard error of 0.25% ($= 1.08\%/\sqrt{18}$). Confidence intervals of two standard errors around an estimate are often used to indicate its accuracy. In this case, our confidence interval for the volatility would be very wide: from around 0.58% to 1.58% per day. Increasing the sample size to 250 would reduce the width of the confidence interval, but it would have the previously mentioned disadvantages.

There is another problem with using Equation (3.1). Suppose there was a large (positive or negative) return 25 days ago and the volatility is calculated using 50 days of data ($m = 50$). For the next 25 days the large return will be included in the sample, and we will tend to get a high volatility as a result. In 26 days, however, the large observation will drop out of the data set, and the volatility will experience a large one-day decline as a result.

Exponential Smoothing

One way of overcoming these problems is to use exponential smoothing. Also referred to as the exponentially weighted moving average (EWMA), this approach was used by RiskMetrics to produce estimated volatilities of a wide range of market variables in the early 1990s.[6]

Equation (3.1) gives equal weight to the squared returns in the sample (i.e., there are m squared returns and the weight given to each squared return is $1/m$). The idea behind EWMA is that while the total of the weights given to all the squared returns must add up to one (as they do in Equation (3.1)), the weights themselves do not have to be equal.

In EWMA, the weight applied to the squared return from k days ago is λ multiplied by the weight applied to the squared return $k - 1$ days ago (where λ is a positive constant that is less than one). Suppose w_0 is the weight applied to the most recent return (i.e., the return on day $n - 1$). The weight for the squared return on day $n - 2$ is λw_0; the weight applied to the squared return on day $n - 3$ is $\lambda^2 w_0$; and so on. The weights are shown in Table 3.2. They decline exponentially as you go further back in time.

The best value of λ is the one that produces the estimate with the lowest error. RiskMetrics found that setting $\lambda = 0.94$ proved to be a good choice across a range of different market variables in the 1990s.

Table 3.2 Weights in the EWMA Method

Day	Squared return	Weight
$n - 1$	r_{n-1}^2	w_0
$n - 2$	r_{n-2}^2	λw_0
$n - 3$	r_{n-3}^2	$\lambda^2 w_0$
$n - 4$	r_{n-4}^2	$\lambda^3 w_0$
..	..	..
..	..	..

[6] RiskMetrics was part of JPMorgan in the early 1990s but was spun off as a separate company in 1998.

Suppose that we choose $\lambda = 0.94$. The weight given to the return 250 business days ago is $0.94^{249}\,w_0$ $(= 0.00000019w_0)$. Because 0.94^{249} is a very small number, the contribution of returns from long ago is also very small.

Suppose that we use K days of data. The total weight applied to all the data items is

$$w_0 + w_0\lambda + w_0\lambda^2 + w_0\lambda^3 + \cdots w_o\lambda^{K-1}$$
$$= w_0(1 + \lambda + \lambda^2 + \lambda^3 + \cdots \lambda^{K-1})$$

As a practical matter, we are likely to initiate EWMA by considering just one or two years of data. But given that data from longer ago carries very little weight, we can assume that data stretching back forever is used.

The total of the weights is then:

$$w_0 \sum_{0}^{\infty} \lambda^k$$

From the formulas for the sum of an infinite geometric series, we get

$$\sum_{i=0}^{\infty} \lambda^k = \frac{1}{1-\lambda}$$

The total of the weights is therefore:

$$w_0\left(\frac{1}{1-\lambda}\right)$$

We require the weights to add up to 1. They will do so when:

$$w_0\left(\frac{1}{1-\lambda}\right) = 1$$

or:

$$w_0 = 1 - \lambda$$

The estimated volatility on day n (which we will denote by σ_n) is given by applying the weights to past squared returns:

$$\sigma_n^2 = w_0 r_{n-1}^2 + w_0\lambda r_{n-2}^2 + w_0\lambda^2 r_{n-3}^2 + \cdots$$

Now consider what happens as we move from day n to day $n-1$. The estimate of the volatility for day $n-1$ is

$$\sigma_{n-1}^2 = w_0 r_{n-2}^2 + w_0\lambda r_{n-3}^2 + w_0\lambda^2 r_{n-4}^2 + \cdots$$

A comparison of this equation with the previous equation shows that

$$\sigma_n^2 = w_0 r_{n-1}^2 + \lambda\sigma_{n-1}^2$$

Substituting $w_0 = 1 - \lambda$, we obtain

$$\sigma_n^2 = (1-\lambda) r_{n-1}^2 + \lambda\sigma_{n-1}^2 \qquad (3.2)$$

This formula provides a very simple way of implementing EWMA. The new estimate of the variance rate on day n is a weighted average of:

- The estimate of the variance rate made for the previous day $(n - 1)$, and
- The most recent observation of the squared return (on day $n - 1$).

The weight given to the most recent variance estimate is λ and the weight given to the new squared return is $1 - \lambda$. Equation (3.2) is sometimes referred to as *adaptive volatility estimation* because prior beliefs about volatility are updated by new data.

An attractive feature of Equation (3.2) is that very little data needs to be stored once EWMA has been implemented for a particular market variable. Indeed, it is only necessary to store the most recent volatility estimate and even the return history does not need to be remembered. When a new return is observed, the volatility estimate is updated and the new return can then be discarded.

As an example of volatility updating using EWMA, suppose today's volatility estimate is 2% per day and we observe a return of −1% today. The current variance rate is $0.02^2 = 0.0004$. The latest squared return is $(-0.01)^2 = 0.0001$. Assuming $\lambda = 0.94$, the new estimate of the variance rate from Equation (3.2) is

$$0.06 \times 0.0001 + 0.94 \times 0.0004 = 0.000382$$

The new volatility estimate is 1.9545% $(= \sqrt{0.000382})$. The estimate has declined from 2% to about 1.95% because the absolute value of the return (i.e., 1%) is less than the 2% current estimate of the volatility.

Determining λ

We now return to the determination of λ. As mentioned earlier, RiskMetrics used $\lambda = 0.94$. Suppose first we use a much higher value of λ (e.g., 0.995). This would make EWMA relatively unresponsive to new data. From Equation (3.2), yesterday's variance rate estimate would be given a weight of 99.5%, while the new squared return estimate would be given a weight of only 0.5%. Even several days of high volatility would not move the estimates much. Setting λ equal to a low number (e.g., 0.5) would lead estimates that overreact to new data. The estimates would themselves be too volatile.

The 0.94 estimate found by RiskMetrics creates a situation where volatility estimates reacted reasonably well to new data (at least in the 1990s). It should be noted that RiskMetrics wanted a λ that could be used for all market variables. If it had

Table 3.3 Scenarios Sorted from the Worst to the Best

Scenario Number	Loss (USD Millions)	Weight	Cumulative Weight
490	7.8	0.0090	0.0090
492	6.5	0.0092	0.0182
2	4.6	0.0001	0.0183
23	4.3	0.0001	0.0184
48	3.9	0.0001	0.0185
367	3.7	0.0026	0.0211
235	3.5	0.0007	0.0218

allowed itself the luxury of using different λs for different market variables, some of the λs used might have differed from 0.94.

One way to determine the optimal λ is to calculate the realized volatility for a particular day using 20 or 30 days of subsequent returns and then search for the value of λ that minimizes the difference between the forecasted volatility and the realized volatility.

Another approach is the maximum likelihood method, which selects a λ to maximize the probability of observed data occurring. As an example, suppose a particular trial value of λ leads to a forecast of a low volatility for a certain day. However, the size of the observed return for that day is large (either positive or negative). The probability of that observation occurring for that value of λ is low, making it less likely that the trial value of λ will prove to be the best estimate.

Historical Simulation

Just as it is possible to apply weights that decline exponentially when calculating volatility, it is also possible to apply weights that decline exponentially when using a historical simulation. The weights must add to one. (This converts historical simulation from a non-parametric approach to a parametric approach.) The method for determining weights is very similar to that for EWMA.

Using exponentially declining weights can be useful when VaR or expected shortfall is calculated from the immediately preceding observations (e.g., observations over the last year.)[7] This is because scenarios calculated from recent data are more relevant than scenarios calculated using data from many days ago. If the economic environment suddenly becomes more stressed (as it did in the second half of 2007), we would want to reflect this in the VaR and expected shortfall. These estimates would be more likely to do so if we applied more weight to scenarios calculated from recent observations.

When weighting schemes are used for historical simulation, scenarios are ranked (as described in Chapter 2) from the largest loss to the smallest. Weights are then accumulated to determine the VaR. Table 3.3 is the same as Table 2.3 from Chapter 2, except that we have added weights and cumulative weights.

The one-day VaR with 99% confidence is now USD 6.5 million (instead of USD 3.9 million as before) because the second loss takes us from the situation where we have seen less than 1% of the (weighted) loss distribution to the situation where we have seen more than 1% of the distribution.

The new estimate of expected shortfall is (in USD million):

$$7.8 \times 0.9 + 6.5 \times 0.1 = 7.67$$

This is because when we are in the 1% tail of the loss distribution, there is a 90% chance that the loss is USD 7.8 million and a 10% chance that it is USD 6.5 million. (Note that the third, fourth, and fifth scenarios have very little weight because they are from very old data.)

The weights in Table 3.3 are approximately consistent with a value for λ equal to 0.99. It turns out that the values of λ that work well for determining weights for historical simulation scenarios are much higher than the values that work well when exponential smoothing is used to weight squared returns.

[7] As noted in Chapter 2, regulators have moved to using stressed VaR and stressed expected shortfall where scenarios are created from a stressed period in the past. It is not appropriate to weight scenarios in this case because there is no reason to suppose that the scenario created from one day during the stressed period is more important than a scenario created from another day during the stressed period.

Alternative Weighting Schemes

In EWMA, the weights applied to historical data decline exponentially as we move back in time. In an alternative method, sometimes referred to as *multivariate density estimation* (MDE), another approach is used to determine weights. With MDE, an analysis is carried out to determine which periods in the past are most similar to the current period. Weights are then assigned to a day's historical data according to how similar that day is to the current day.

Consider interest rates. The volatility of an interest rate tends to vary according to interest rate levels. Specifically, volatility (measured as the standard deviation of percentage daily changes) tends to decrease as rates increase. In calculating the volatility of an interest rate, it can make sense to give more weight to data from periods where the level of interest rates is similar to current interest rates levels and allow the weight for a day's data to taper off as the difference between the interest rate levels for that day and the current day increases.

Sometimes several state variables (referred to as conditioning variables) are used to determine how similar one period is to another. For example, we could use GDP growth in conjunction with interest rate levels as conditioning variables for interest rate volatility. When there are several conditioning variables (X_1., X_2, . . . , X_n) we determine the similarity between today and a previous day by calculating a metric such as:

$$\sum_{i=1}^{n} a_i(\hat{X}_i - X_i^*)^2$$

where X_i^* is the value of X_i today, $\hat{X}_i$ is the value of X_i on the day from history being considered, and a_i is a constant reflecting the importance of the *i*th variable. The more similar a historical day is to the current day, the smaller this metric becomes.

3.5 GARCH

The GARCH model, developed by Robert Engel and Tim Bollerslev, can be regarded as an extension of EWMA.[8] In EWMA, Equation (3.2) shows that we give some weight to the most recent variance rate estimate and some weight to the latest squared return. In GARCH (1,1), we also give some weight to a long run average variance rate. The updated formula for the variance rate is

$$\sigma_n^2 = \alpha r_{n-1}^2 + \beta\sigma_{n-1}^2 + \gamma V_L \qquad (3.3)$$

[8] GARCH is short for generalized autoregressive conditional heteroscedasticity.

Here, V_L is the long run average variance rate. The parameters α, β, and γ are the weights given to the most recent squared return, the previous variance rate estimate, and the long run average variance rate (respectively). Because the weights must sum to one:

$$\alpha + \beta \leq 1, \text{ and}$$

$$\gamma = 1 - \alpha - \beta$$

where α and β positive, and the unconditional variance has been normalized to one.

Comparing Equations (3.2) and (3.3), we see EWMA is a particular case of GARCH (1,1) where $\gamma = 0$, $\alpha = 1 - \lambda$ and $\beta = \lambda$. Both GARCH (1,1) and EWMA can be classified as *first-order autoregressive models* because the value forecast for the variable (variance rate in this case) depends on the immediately preceding value of the variable. First order autoregressive models are also referred to as AR(1) models.

The (1,1) in GARCH (1,1) indicates weight is given to one (most recently observed) squared return and one (most recent) variance rate estimate. In the more general GARCH (*p*,*q*), weight is given to the most recent *p* squared returns and the most recent *q* variance rate estimates. GARCH (1,1) is by far the most widely used version of GARCH.

The weights in a GARCH (1,1) model decline exponentially in the same way that the weights in EWMA decline exponentially. We can see this by noting from Equation (3.3):

$$\sigma_{n-1}^2 = \alpha r_{n-2}^2 + \beta\sigma_{n-2}^2 + \gamma V_L$$

$$\sigma_{n-2}^2 = \alpha r_{n-3}^2 + \beta\sigma_{n-3}^2 + \gamma V_L$$

$$\sigma_{n-3}^2 = \alpha r_{n-4}^2 + \beta\sigma_{n-4}^2 + \gamma V_L$$

and using these equations to substitute successively for $\sigma_{n-1}^2, \sigma_{n-2}^2, \sigma_{n-3}^2, \ldots$. starting with Equation (3.3).

Equation (3.3) has parameters: α, β, γ, and V_L. It is usual to define

$$\omega = \gamma V_L$$

so that there are only three parameters and:

$$\sigma_n^2 = \omega + \alpha r_{n-1}^2 + \beta\sigma_{n-1}^2 \qquad (3.4)$$

Because $\alpha + \beta + \gamma = 1$:

$$V_L = \frac{\omega}{1 - \alpha - \beta}$$

As an example of GARCH (1,1) calculations, suppose $\omega = 0.000003$, $\alpha = 0.12$, and $\beta = 0.87$ so that Equation (3.4) is

$$\sigma_n^2 = 0.000003 + 0.12 r_{n-1}^2 + 0.87\sigma_{n-1}^2$$

In this case $\gamma = 1 - \alpha - \beta = 0.01$ and $V_L = 0.000003/0.01 = 0.0003$. The long-run average variance rate is 0.0003. This corresponds to a volatility of 1.732% (= $\sqrt{0.0003} = 0.01732$). Suppose the estimated current volatility σ_{n-1} is 2% per day and the new return r_{n-1} is −3%. The new variance rate estimate is

$$0.000003 + 0.12 \times (-0.03)^2 + 0.87 \times 0.02^2 = 0.000459$$

This corresponds to a volatility of 2.14% (= $\sqrt{0.000459}$). The volatility has increased because the latest observation on the return is greater than the current volatility. However, it has also been pulled down slightly because the long-run average volatility is only 1.732%.

Mean Reversion

The difference between EWMA and GARCH (1,1) is that in GARCH (1,1), the V_L term provides a "pull" toward the long-run average mean. In EWMA, however, there is no pull because this term is not included. Figure 3.3 illustrates the pull. When volatility is above the long run average mean, it tends to be pulled down toward it. When the volatility is below the long run average mean, it tends to get pulled up toward it. Of course, this is just a tendency. The observed returns will lead to ups and downs superimposed on the average path given by the V_L term.

This "pull" is referred to as *mean reversion*. GARCH(1,1) incorporates mean reversion whereas EWMA does not. It is important to note that mean reversion is a reasonable property for some market variables, but not for others. If the market variable is the price of something that can be traded, it should not exhibit a predictable mean reversion because otherwise there would be a market inefficiency which can be exploited for profit.

A volatility is not the price of something that can be traded. There is, therefore, nothing to stop volatility exhibiting mean reversion. Indeed, there is every reason to suppose that it does. For example, when volatility becomes unusually high or low, we do not expect that to last forever. It will eventually revert to normal levels. The same argument can be used for an interest rate. An interest rate is not the price of a traded security.[9] Interest rates can therefore follow mean reversion and in practice appear to do so.

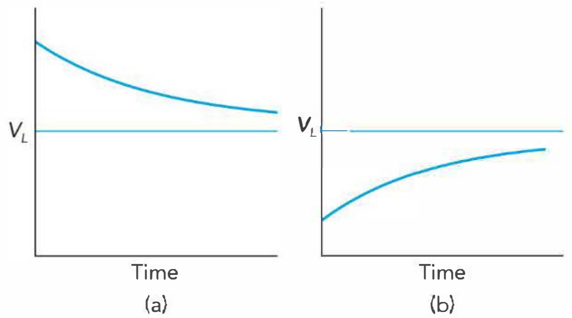


Figure 3.3 **Shown is the effect of mean reversion on the future expected variance rate. In (a) the initial variance rate is above the long-run average; in (b) it is below the long-run average.**

Long Horizon Volatility

So far, we have focused on estimating the one-day volatility. If we assume a conditional normal model, the one-day volatility can be used to provide a normal distribution for the change in a market variable. This, in turn, can be used as an input to the delta-normal model for determining the one-day VaR.

Often, we are interested in what might happen over a longer period of time than one day. If volatility is constant, it is reasonable to assume that the variance rate over T days is T times the variance rate over one day. This means that the volatility over T days is $\sqrt{T}$ times the volatility over one day. This square root rule corresponds to a well-known rule of thumb that uncertainty increases with the square root of time. This was used in the results considered in earlier chapters, where value at risk with a certain confidence level over T days is assumed to be $\sqrt{T}$ multiplied by the VaR over one day.[10]

Mean reversion suggests a possible improvement to the square root rule. If the current daily volatility is high, we expect it to decline. That means the square root rule based on the current daily volatility overstates VaR. Similarly, if the current daily volatility is low, we expect it to increase and so the square root rule will tend to understate VaR. The same phenomenon applies to expected shortfall. It can be shown that under GARCH (1,1), the expected variance rate on day t is given by

$$\sigma_{n+t}^2 = V_L + (\alpha + \beta)^t(\sigma_n^2 - V_L)$$

We can use this formula to calculate the average variance rate over the next T days. The expected volatility can be estimated as the square root of this average variance rate.

Suppose that our estimate of the current daily variance is 0.00010 (corresponding to a volatility of 1% per day). Due to mean reversion, however, we expect the average daily variance rate over the next 25 days to be 0.00018. It makes sense then to base VaR and expected shortfall estimates for a 25-day time horizon on volatility over the 25 days. This is equal to 6.7% (= $\sqrt{25} \times \sqrt{0.00018} = 0.067$).

[9] A bond price is a function of interest rates and is the price of something that can be traded but the interest rate itself is not. Similarly we can trade instruments that depend on volatility, but we cannot trade something whose price always equals a volatility.

[10] This result is exact in the situation where changes in successive days have identical normal distributions with mean zero. Otherwise it is an approximation.

3.6 IMPLIED VOLATILITY

No discussion of volatility would be complete without mentioning implied volatility. The implied volatility is the volatility implied from option prices. We will discuss this in some detail in later chapters. At this stage we note the value of a call or put option depends on volatility. As volatilities increase, the value of the option increases. This is what enables a volatility to be implied from an option price. Implied volatilities are forward looking, whereas the volatilities calculated from historical data (e.g., EWMA, GARCH, or some other method) are backward looking. When market participants are pricing options, they are thinking about what volatility will be in the future, not what it has been in the past.

Evidence indicates that implied volatilities give better estimates of realized volatility than historical volatilities. The implied volatility of an option is usually expressed as a volatility per year. It can be converted to a volatility per day by dividing by the square root of 252 (252 is an estimate of the number of trading days in a year). Thus, a volatility of 20% per year corresponds to a volatility of 1.26% ($= 20\%/\sqrt{252}$) per day.

The implied volatility of a one-month option gives an indication of the average volatility expected over the next month, the implied volatility of a three-month option gives an indication of what volatility is expected to be over the next three months, and so on. Because volatility exhibits mean reversion, we do not expect implied volatilities to be the same for options of all maturities.

Options are not actively traded on all assets. This means reliable implied volatilities for an asset are sometimes not available. Nevertheless, risk managers should monitor implied volatilities as well as volatilities calculated from historical data whenever possible.

One closely monitored implied volatility index is the VIX. This is an index of the implied volatilities of 30-day options on the S&P 500. Figure 3.4 shows a history of the VIX Index. Typical values for the index are in the 10 to 20 range (indicating a 30-day volatility for the S&P 500 of 10% to 20% per year). During October 2008 (in the midst of the crisis), the VIX index reached 80 (indicating an implied volatility of 80% per year and 5% per day) on two occasions. The index spiked again in March 2020 as a result of the coronavirus, which led to huge movements in the S&P 500.

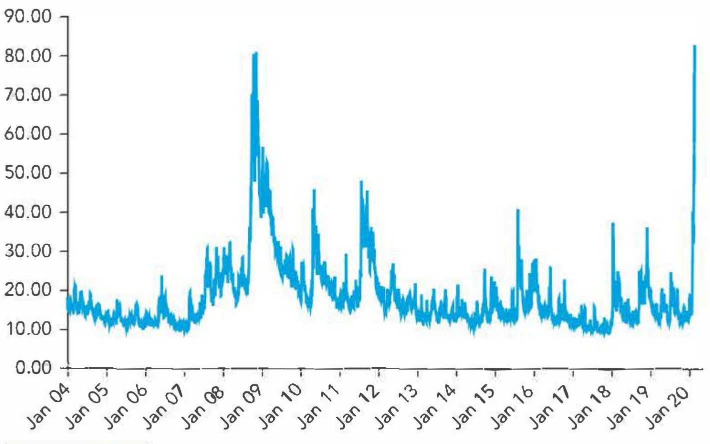


Figure 3.4 **History of the VIX index: January 2004 to March 2020.**

3.7 CORRELATION

It is important for risk managers to monitor correlations as well as volatilities. When the delta-normal model is used to calculate VaR or expected shortfall for a linear portfolio, correlations between daily asset returns (as well the volatilities of daily returns) are needed.

To update correlation estimates, we can use rules similar to those we use for updating volatility estimates. Just as updating rules for volatility work with variances, the updating rules for correlations work with covariances. If the mean daily returns are assumed to be zero, the covariance of the returns between two variables is the expectation of the product of the returns. The EWMA model for updating the covariance between return X and return Y is

$$\text{cov}_n = \lambda \text{cov}_{n-1} + (1 - \lambda)x_{n-1}y_{n-1} \tag{3.5}$$

where cov_n is the covariance estimated for day n and x_n and y_n are values of X and Y on day n (respectively).

The correlation between two variables is their covariance divided by the product of their standard deviations. If EWMA has been used to estimate the standard deviations of the returns, we can estimate the coefficient of correlation. For consistency we should use the same value of λ for updating both variance rates using Equation (3.2) and covariances using Equation (3.5).

Suppose the volatility of X and Y are estimated for day $n - 1$ as 1% and 2% per day (respectively) while the coefficient of correlation has been estimated as 0.2. The covariance is the coefficient of correlation multiplied by the product of the standard deviations:

$$\text{cov}_{n-1} = 0.2 \times 0.01 \times 0.02 = 0.00004$$

Suppose that the returns observed on day $n - 1$ for X and Y are both 2% and we use a value for λ of 0.94. The covariance is updated as follows:

$$\text{cov}_n = 0.94 \times 0.00004 + 0.06 \times 0.02 \times 0.02 = 0.0000616$$

With the same value of λ, the volatilities of X and Y are updated to 1.086% and 2% (respectively) and the new coefficient of correlation is

$$\frac{0.0000616}{0.01086 \times 0.02} = 0.28$$

GARCH (1,1) can also be used to update a covariance. However, using GARCH to update multiple covariances in a consistent way is quite complex.

SUMMARY

Volatilities vary through time. Thus, a model where returns are conditionally normal is better than a model where returns are unconditionally normal, and it is, therefore, important for risk managers to keep track of volatilities. The usual formulas for calculating variances from historical data can be used, but these have a disadvantage in that they may not be responsive enough to a changing volatility environment. Also, values may be unduly influenced by whether a particular extreme observation is included in the historical sample.

The exponentially weighted moving average (EWMA) method is one approach for monitoring volatility. It involves giving weights to observations that decrease exponentially as we move back in time. GARCH (1,1) is a more sophisticated alternative to EWMA that gives some weight to a long-run average variance rate in the updating formula. This recognizes that volatilities exhibit mean reversion and provides a better estimate of volatility over periods longer than one day.

Implied volatility is the volatility implied by option prices observed in the market. It is a forward-looking estimate of volatility and is often found to be more accurate than an estimate produced from historical data.

Correlations (like volatilities) vary through time. The EWMA tool can be used to monitor covariances in the same way as it is used to monitor variances. The current estimate of a correlation between two variables can be calculated by dividing the current estimate of a covariance by the product of the standard deviations of the variables.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

QUESTIONS

Short Concept Questions

3.1 What mixtures of normal distributions give (a) a distribution with fat tails and (b) a non-symmetrical distribution?

3.2 Explain the difference between conditional and unconditional normality.

3.3 What is meant by regime switching?

3.4 What is the assumption underlying the exponentially weighted moving average (EWMA) model for estimating volatility?

3.5 How does GARCH (1,1) differ from EWMA?

3.6 How is the parameter λ defined in EWMA? What value did RiskMetrics use for λ? What is the possible range of values for lambda? What are the effects of choosing lambda too small or too high?

3.7 What are the adaptive volatility estimation formulas for (a) EWMA and (b) GARCH (1,1)?

3.8 If daily volatility is currently higher that the long-run average volatility, would you expect the square root rule to overstate or understate the volatility over a long-horizon volatility?

3.9 Does implied volatility or historical volatility give a better estimate of future volatility?

3.10 What is the relationship between covariance and coefficient of correlation?

Practice Questions

3.11 A stock price (USD) is 50 and the volatility is 2% per day. Estimate a 95% confidence interval for the stock price at the end of one day.

3.12 A company uses EWMA to estimate volatility day-by-day. What would be the general effect of changing the volatility parameter from 0.94 to (a) 0.84 and (b) 0.98?

3.13 Suppose that the observations on a stock price (in USD) at the close of trading on 15 consecutive days are 40.2, 40.0, 41.1, 41.0, 40.2, 42.3, 43.1, 43.4, 42.9, 41.8, 43.7, 44.3, 44.4, 44.8, and 45.1.

Estimate the daily volatility using Equation (3.1).

3.14 Suppose that the price of an asset at the close of trading yesterday was USD 20 and its volatility was estimated as 1.4% per day. The price at the close of trading today is USD 19. What is the new volatility estimate using the EWMA with a λ of 0.9?

3.15 If $\omega = 0.000002$, $\alpha = 0.04$, and $\beta = 0.94$ in a GARCH model, what is the long-run average variance rate? What volatility does this correspond to?

3.16 Repeat Question 3.14 using the GARCH (1,1) model in Question 3.15.

3.17 Suppose that the current volatility estimate is 3% per day and the long-run average volatility estimate is 2% per day. What are the volatility estimates in ten days and 100 days in a GARCH (1,1) model where $\omega = 0.000002$, $\alpha = 0.04$, and $\beta = 0.94$?

3.18 Delete scenario number 490 from Table 3.3 so that the worst scenario is scenario 492 with a loss of USD 6.5. Assume there are still 500 scenarios. What is the new VaR?

3.19 In Question 3.18, what is the new 99% expected shortfall estimate?

3.20 Suppose that the price of asset X at the close of trading yesterday was USD 40 and its volatility was then estimated as 1.5% per day. Suppose further that the price of asset Y at the close of trading yesterday was USD 10 and its volatility was then estimated as 1.7% per day. The price of X and Y at the close of trading today are USD 38 and USD 10.1, respectively. The correlation between X and Y was estimated as 0.4 at the close of trading yesterday. Update the volatility of X and Y and the correlation between X and Y using the EWMA model with λ equal to 0.95.

ANSWERS

Short Concept Questions

3.1 When normal distributions with the same mean and different standard deviations are mixed, we get fat tails. When both their means and standard deviations are different, we get a non-symmetrical distribution.

3.2 Conditional normality means that the distribution of daily returns is normal conditional on the volatility being known. Unconditional normality means that the distribution of daily returns always has the same normal distribution.

3.3 Regime switching is a sudden unexpected change in volatility or another economic variable.

3.4 The weights given to squared returns decline exponentially.

3.5 GARCH (1,1) gives some weight to the long-run average variance rate. This induces mean reversion.

3.6 As we move back in time, each weight is λ multiplied the previous one (where λ is a number between zero and one). RiskMetrics used $\lambda = 0.94$. If λ to high, the resulting estimate will not be sufficiently responsive to changing conditions. If λ is too low, the resulting estimate will be too volatile.

3.7 For (a) see Equation (3.2) and for (b) see Equations (3.3) or (3.4).

3.8 The square root rule will overstate the long-horizon volatility.

3.9 Implied volatility gives the best estimate

3.10 The covariance between two variables is the correlation times the product of the standard deviations.

Solved Problems

3.11 A one standard deviation move is USD 1. The 95% confidence interval is −USD 1.96 to +USD 1.96. The 95% confidence interval at the end of one day is, therefore, USD 48.04 to 51.96.

3.12 (a) The day-to-day changes in the volatility estimates would be more variable; (b) the day-to-day changes in the volatility estimates would be less variable.

3.13 The 14 daily returns are −0.498%, 2.75%, −2.43%, ... The average of the squared returns is 0.000534, and the volatility estimate is the square root of this or 2.31%.

3.14 The new return is $-1/20 = -0.05$. The new variance rate estimate is

$$0.9 \times 0.014^2 + 0.1 \times (-0.05)^2 = 0.000426$$

The new volatility is the square root of this or 2.06%.

3.15 The long-run average variance rate is $0.000002/(1 - 0.04 - 094) = 0.0001$. This corresponds to a volatility of 1% per day.

3.16 The new variance rate is found using Equation (3.4):

$$0.000002 + 0.04 \times (-0.05)^2 + 0.94 \times 0.014^2 = 0.000286$$

The new volatility is the square root of this, which is 1.69%.

3.17 The expected variance rate in ten days is $0.02^2 + (0.04 + 0.94)^{10} (0.03^2 - 0.02^2) = 0.000809$, which corresponds to a volatility of 2.84%. The expected variance rate in 100 days is $0.02^2 + (0.04 + 0.94)^{100} (0.03^2 - 0.02^2) = 0.000466$, which corresponds to a volatility of 2.16%.

3.18 The VaR becomes USD 3.7 million.

3.19 The expected shortfall becomes

$$\frac{0.0092 \times 6.5 + 0.0001 \times 4.6 + 0.0001 \times 4.3 + 0.0001 \times 3.9 + 0.0005 \times 3.7}{0.01}$$

$$= 6.29$$

3.20 The new estimates of the variance rate of X is

$$0.95 \times 0.015^2 + 0.05 \times (-2/40)^2 = 0.000339$$

which corresponds to a new volatility of 1.84%. The new volatility for Y is 1.67%. The covariance yesterday was $0.015 \times 0.017 \times 0.4 = 0.000102$. The new covariance is

$$0.95 \times 0.000102 + 0.05 \times (-2/40) \times (0.1/10) = 0.0000719$$

The new correlation is $0.0000719/(0.0184 \times 0.0167) = 0.23$.

4

External and Internal Credit Ratings

Learning Objectives

After completing this reading, you should be able to:

- Describe external rating scales, the rating process, and the link between ratings and default.
- Define conditional and unconditional default probabilities and explain the distinction between the two.
- Define hazard rate and calculate the unconditional default probability of a credit asset using hazard rate.
- Define recovery rate and calculate the expected loss from a loan.
- Explain and compare the through-the-cycle and point-in-time ratings approaches.
- Describe alternative methods to credit ratings produced by rating agencies.
- Compare external and internal ratings approaches.
- Describe, calculate, and interpret a rating transition matrix and explain its uses.
- Describe the relationships between changes in credit ratings and changes in stock prices, bond prices, and credit default swap spreads.
- Explain historical failures and potential challenges to the use of credit ratings in making investment decisions.

The first three chapters were primarily concerned with market risk. In this chapter, we start to discuss credit risk.

Rating agencies are an important external source of credit risk data. The most well-known credit rating agencies are Moody's, Standard and Poor's (S&P), and Fitch. All three are based in the United States (although they have offices in other countries). There are also many smaller rating agencies (e.g., DBRS) throughout the world.

These agencies provide independent opinions on credit risk, based on specified criteria. They try to make their ratings consistent across regions, industries, and time (with mixed success). Traditionally, the main business of these agencies has been the rating of bonds and money market instruments issued by corporations and governments. Rating agencies have been doing this for more than 100 years and their track record has generally been quite good.

One notable exception is the case of structured products (i.e., investment products formed from portfolios of other products). In the run up to the credit crisis of 2007–2008, rating agencies were increasingly involved in the rating of structured products built around subprime mortgages. Indeed, it is considered that rating agencies contributed to the 2007–2008 crisis by giving excessively high ratings to many structured subprime products that subsequently defaulted.

The role of rating agencies in the credit crisis has had a number of consequences in the United States. The Dodd-Frank Act now requires rating agencies to make the assumptions and methodologies underlying their ratings more transparent. It has also increased the potential legal liability of rating agencies.[1] Meanwhile, the Office of Credit Ratings was created at the Securities and Exchange Commission to provide oversight of rating agencies.

Credit ratings have been widely used by regulators. As early as the 1930s, banks were prohibited from investing in firms with low credit ratings. Today, the Basel Committee uses credit ratings when determining credit risk capital for banks. Since the 2007–2008 crisis, however, the United States has indicated it does not want to do this (presumably because it no longer trusts the ratings). For some capital determinations, two alternative calculations have been specified by the Basel Committee: one for countries that are prepared to use external ratings and one for countries that are not.

The reputations of rating agencies have suffered since the crisis because of their poor performance in rating structured products. However, their performance in rating bonds and money market instruments has been generally good, and many risk managers still rely on these ratings.

External ratings can be of limited use to risk managers because rating agencies (usually) only look at firms with publicly traded bonds or money market instruments. Firms that fund themselves with bank borrowing and do not issue debt instruments are often not assessed by rating agencies. Because many of their borrowers do not have external ratings, banks have developed their own internal rating systems to help them make lending decisions. These rating systems are structured like those of the major rating agencies.

In this chapter, we first describe the rating scales and the rating process. We review data produced by the rating agencies on default probabilities and rating transitions. We then discuss the consistency of ratings across economic cycles, industries, countries, and time. We define hazard rates as well as consider the determinants of recovery rates and credit spreads. We review research on the impact of ratings on bond prices, stock prices, and credit default swap spreads, and contrast the way ratings are produced for structured products with the way they are produced for bonds and money market instruments.

4.1 RATING SCALES

Credit ratings are designed to answer the question: "How likely is an entity to default on its obligations?"[2] An external credit rating is usually an attribute of an instrument issued by an entity (rather than of the entity itself).[3] A bond rating can depend on various factors (e.g., collateral, the term of the instrument, and so on), but often an agency will give all instruments issued by an entity the same rating. As a result, bond ratings are often assumed to be attributes of the entity rather than of the bond itself. For example, we might say that firm X has a rating of BBB when in fact the rating refers to specific bonds issued by firm X.

Ratings agencies use different scales for bonds and money market instruments. The ratings for bonds are termed *long-term ratings*, whereas those for money-market instruments are termed *short-term ratings*. Money market instruments usually last one year or less and provide their entire return in the form of a final

[1] Under the First Amendment of the U.S. Constitution, rating agencies have been protected because they have made it clear that all they were doing was offering an opinion.

[2] In their documentation, S&P and Fitch state that it is probability of default that is being measured. Moody's states that its ratings are designed to measure expected loss from a default. This is probability of default times loss given default. In theory, if a potential default would give rise to little or no loss, Moody's should give a higher rating than S&P.

[3] However, the agencies do sometimes provide issuer ratings as well issue-specific ratings.

payment. In contrast, bonds provide periodic coupon payments to the holder. We now consider the long-term and short-term scales used by rating agencies.

Long-Term Ratings

The highest bond rating assigned by Moody's is Aaa. Bonds with this rating are considered to have almost no chance of defaulting. The next highest rating is Aa. After that, the following ratings (in order of decreasing creditworthiness) are A, Baa, Ba, B, Caa, Ca, and C. The S&P ratings corresponding to those of Moody's are AAA, AA, A, BBB, BB, B, CCC, CC, and C, respectively. The rating of D is used for firms already in default.

Modifiers are used to create finer rating measures. For example, 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 their two lowest rating categories. Fitch's rating categories are like those of S&P.

It is usually assumed the ratings assigned by the different agencies are equivalent. 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 *investment grade*. Those with ratings below BBB− (Baa3) are termed *non-investment grade*, *speculative* grade, or *junk bonds*.

Short-Term Ratings

The rating agencies have different ways of rating money market instruments. Moody's uses P-1, P-2, and P-3 as its three prime rating categories. Instruments rated P-1 are considered to have

Table 4.1 Bond Rating Scales and S&P Definitions

Moody's	S&P and Fitch	S&P Definition
Aaa	AAA	An obligation rated 'AAA' has the highest rating assigned by S&P Global Ratings. The obligor's capacity to meet its financial commitment on the obligation is extremely strong.
Aa	AA	An obligation rated 'AA' differs from the highest-rated obligations only to a small degree. The obligor's capacity to meet its financial commitment on the obligation is very strong.
A	A	An obligation rated 'A' is somewhat more susceptible to the adverse effects of changes in circumstances and economic conditions than obligations in higher-rated categories. However, the obligor's capacity to meet its financial commitment on the obligation is still strong.
Baa	BBB	An obligation rated 'BBB' exhibits adequate protection parameters. However, adverse economic conditions or changing circumstances are more likely to lead to a weakened capacity of the obligor to meet its financial commitment on the obligation.
Ba	BB	An obligation rated 'BB' is less vulnerable to nonpayment than other speculative issues. However, it faces major ongoing uncertainties or exposure to adverse business, financial, or economic conditions which could lead to the obligor's inadequate capacity to meet its financial commitment on the obligation.
B	B	An obligation rated 'B' is more vulnerable to nonpayment than obligations rated 'BB', but the obligor currently has the capacity to meet its financial commitment on the obligation. Adverse business, financial, or economic conditions will likely impair the obligor's capacity or willingness to meet its financial commitment on the obligation.
Caa	CCC	An obligation rated 'CCC' is currently vulnerable to nonpayment, and is dependent upon favorable business, financial, and economic conditions for the obligor to meet its financial commitment on the obligation. In the event of adverse business, financial, or economic conditions, the obligor is not likely to have the capacity to meet its financial commitment on the obligation.
Ca	CC	An obligation rated 'CC' is currently highly vulnerable to nonpayment. The 'CC' rating is used when a default has not yet occurred, but S&P Global Ratings expects default to be a virtual certainty, regardless of the anticipated time to default.
C	C	An obligation rated 'C' is currently highly vulnerable to nonpayment, and the obligation is expected to have lower relative seniority or lower ultimate recovery compared to obligations that are rated higher.

a superior ability to repay short-term obligations, instruments rated P-2 have a strong ability to do so, and instruments rated P-3 have an acceptable ability to do so. These can be thought of as the investment grade ratings. The rating NP denotes non-prime and can be thought of as a non-investment grade rating. The S&P rating category corresponding to P-1 is divided into two: A-1+ and A-1. The categories equivalent to P-2 and P-3 are A-2 and A-3 (respectively). Meanwhile, S&P has three lower rating categories: B, C, and D. Fitch subdivides its ratings in a similar way to S&P, with its ratings being F1+, F1, F2, F3, B, C, and D. This is summarized together with the definitions used by S&P in Table 4.2.

Table 4.2 Short-Term Rating Scales and S&P Definitions

Moody's	S&P	Fitch	S&P Definition
P-1	A-1	F1	A short-term obligation rated 'A-1' is rated in the highest category by S&P Global Ratings. The obligor's capacity to meet its financial commitment on the obligation is strong. Within this category, certain obligations are designated with a plus sign (+). This indicates that the obligor's capacity to meet its financial commitment on these obligations is extremely strong.
P-2	A-2	F2	A short-term obligation rated 'A-2' is somewhat more susceptible to the adverse effects of changes in circumstances and economic conditions than obligations in higher rating categories. However, the obligor's capacity to meet its financial commitment on the obligation is satisfactory.
P-3	A-3	F3	A short-term obligation rated 'A-3' exhibits adequate protection parameters. However, adverse economic conditions or changing circumstances are more likely to lead to a weakened capacity of the obligor to meet its financial commitment on the obligation.
NP	B	B	A short-term obligation rated 'B' is regarded as vulnerable and has significant speculative characteristics. The obligor currently has the capacity to meet its financial commitments; however, it faces major ongoing uncertainties which could lead to the obligor's inadequate capacity to meet its financial commitments.
	C	C	A short-term obligation rated 'C' is currently vulnerable to nonpayment and is dependent upon favorable business, financial, and economic conditions for the obligor to meet its financial commitment on the obligation.
	D	D	A short-term obligation rated 'D' is in default or in breach of an imputed promise. For non-hybrid capital instruments, the 'D' rating category is used when payments on an obligation are not made on the date due, unless S&P Global Ratings believes that such payments will be made within any stated grace period. However, any stated grace period longer than five business days will be treated as five business days. The 'D' rating also will be used upon the filing of a bankruptcy petition or the taking of a similar action and where default on an obligation is a virtual certainty, for example due to automatic stay provisions. An obligation's rating is lowered to 'D' if it is subject to a distressed exchange offer.

4.2 HISTORICAL PERFORMANCE

One way of testing the performance of these ratings is to look at how firms with a specific rating subsequently performed. Table 4.3 shows the cumulative average default rates produced by S&P for bonds for the period from 1981 to 2018 (similar tables are produced by Moody's and Fitch). It shows the probability that an issuer that starts with a certain rating will default within one year, within two years, within three years, and so on. As already mentioned, investment grade issuers are those with a rating of BBB or above. Speculative grade firms are those with ratings of BB or below.

It shows, for example, that an issuer with a rating of A has a 0.06% chance (6 in 10,000) of defaulting within one year, 0.14% chance of defaulting within two years, 0.23% chance of defaulting within three years, and so on.

We can calculate the probability of a bond defaulting during a future year by subtracting successive numbers in Table 4.3. For example, the probability of a B-rated bond defaulting during the fifth year is 2.38% (= 17.33% − 14.95%). The results of this calculation are in Table 4.4 and we will refer to them as unconditional default probabilities.

Another way of expressing the probability of default during a given year is as a conditional default probability. In this case, we are asking the question; "If the firm survives to the end of year n, what is the probability that it will default during year $n + 1$?" Consider again the probability a bond rated B will default during the fifth year. The probability it will survive to the end of the fourth

Table 4.3 Global Corporate Average Cumulative Default Rates (1981–2018). C35 Calculated by S&P from its Ratings

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
AAA	0.00	0.03	0.13	0.24	0.35	0.45	0.51	0.59	0.65	0.70	0.73	0.76	0.79	0.85	0.92
AA	0.02	0.06	0.12	0.22	0.32	0.42	0.51	0.59	0.66	0.73	0.80	0.86	0.92	0.98	1.04
A	0.06	0.14	0.23	0.35	0.49	0.63	0.81	0.96	1.12	1.28	1.43	1.57	1.71	1.83	1.98
BBB	0.17	0.46	0.80	1.22	1.64	2.05	2.41	2.76	3.11	3.44	3.79	4.06	4.32	4.59	4.87
BB	0.65	2.01	3.63	5.25	6.78	8.17	9.36	10.43	11.38	12.22	12.92	13.56	14.13	14.63	15.17
B	3.44	7.94	11.86	14.95	17.33	19.26	20.83	22.07	23.18	24.21	25.08	25.73	26.31	26.87	27.43
CCC/C	26.89	36.27	41.13	43.94	46.06	46.99	48.20	49.04	49.80	50.44	50.96	51.51	52.16	52.72	52.80
Investment Grade	0.09	0.25	0.43	0.66	0.90	1.14	1.36	1.56	1.77	1.96	2.16	2.32	2.48	2.63	2.80
Speculative Grade	3.66	7.13	10.12	12.56	14.55	16.18	17.55	18.69	19.70	20.62	21.39	22.02	22.60	23.13	23.65
All Rated	1.48	2.91	4.16	5.21	6.08	6.82	7.44	7.97	8.44	8.88	9.26	9.58	9.87	10.13	10.41

Source: *2018 Annual Global Corporate Default And Rating Transition Study* © 2019, reproduced with permission of S&P Global Market Intelligence LLC and Standard & Poor's Financial Services LLC.

Table 4.4 Unconditional Percentage Default Probabilities Calculated from Table 4.3

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
AAA	0.00	0.03	0.10	0.11	0.11	0.10	0.06	0.08	0.06	0.05	0.03	0.03	0.03	0.06	0.07
AA	0.02	0.04	0.06	0.10	0.10	0.10	0.09	0.08	0.07	0.07	0.07	0.06	0.06	0.06	0.06
A	0.06	0.08	0.09	0.12	0.14	0.14	0.18	0.15	0.16	0.16	0.15	0.14	0.14	0.12	0.15
BBB	0.17	0.29	0.34	0.42	0.42	0.41	0.36	0.35	0.35	0.33	0.35	0.27	0.26	0.27	0.28
BB	0.65	1.36	1.62	1.62	1.53	1.39	1.19	1.07	0.95	0.84	0.70	0.64	0.57	0.50	0.54
B	3.44	4.50	3.92	3.09	2.38	1.93	1.57	1.24	1.11	1.03	0.87	0.65	0.58	0.56	0.56
CCC/C	26.89	9.38	4.86	2.81	2.12	0.93	1.21	0.84	0.76	0.64	0.52	0.55	0.65	0.56	0.08
Investment Grade	0.09	0.16	0.18	0.23	0.24	0.24	0.22	0.20	0.21	0.19	0.20	0.16	0.16	0.15	0.17
Speculative Grade	3.66	3.47	2.99	2.44	1.99	1.63	1.37	1.14	1.01	0.92	0.77	0.63	0.58	0.53	0.52
All Rated	1.48	1.43	1.25	1.05	0.87	0.74	0.62	0.53	0.47	0.44	0.38	0.32	0.29	0.26	0.28

Source: *2018 Annual Global Corporate Default And Rating Transition Study* © 2019, reproduced with permission of S&P Global Market Intelligence LLC and Standard & Poor's Financial Services LLC.

year is 85.05% (= 100% − 14.95%) based the probabilities from Table 4.3. We have calculated the unconditional probability that the bond will default during the fifth year as 2.38% (or 0.0238). The probability that it will default during the fifth year, conditional on no earlier default, is therefore 2.80% (= 0.0238/0.8505). Table 4.5 shows the results of these conditional default probability calculations for all rating categories and all future years.

Conditional and unconditional default probabilities are simply different ways of expressing the same data. An analogy here might be useful. The unconditional probability than a man will die when aged 99 is quite small. However, the probability that he will die aged 99 conditional that he lives to age 98 is much larger. Both probabilities can be calculated from mortality tables.

Table 4.5 Probability of Default Conditional on No Earlier Default from Table 4.3

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
AAA	0.00	0.03	0.10	0.11	0.11	0.10	0.06	0.08	0.06	0.05	0.03	0.03	0.03	0.06	0.07
AA	0.02	0.04	0.06	0.10	0.10	0.10	0.09	0.08	0.07	0.07	0.07	0.06	0.06	0.06	0.06
A	0.06	0.08	0.09	0.12	0.14	0.14	0.18	0.15	0.16	0.16	0.15	0.14	0.14	0.12	0.15
BBB	0.17	0.29	0.34	0.42	0.43	0.42	0.37	0.36	0.36	0.34	0.36	0.28	0.27	0.28	0.29
BB	0.65	1.37	1.65	1.68	1.61	1.49	1.30	1.18	1.06	0.95	0.80	0.73	0.66	0.58	0.63
B	3.44	4.66	4.26	3.51	2.80	2.33	1.94	1.57	1.42	1.34	1.15	0.87	0.78	0.76	0.77
CCC/C	26.89	12.83	7.63	4.77	3.78	1.72	2.28	1.62	1.49	1.27	1.05	1.12	1.34	1.17	0.17
Investment Grade	0.09	0.16	0.18	0.23	0.24	0.24	0.22	0.20	0.21	0.19	0.20	0.16	0.16	0.15	0.17
Speculative Grade	3.66	3.60	3.22	2.71	2.28	1.91	1.63	1.38	1.24	1.15	0.97	0.80	0.74	0.68	0.68
All Rated	1.48	1.45	1.29	1.10	0.92	0.79	0.67	0.57	0.51	0.48	0.42	0.35	0.32	0.29	0.31

Tables 4.4 and 4.5 show that for the investment grade bonds, the probability of default per year is an increasing function of time for the first four years. This is because the bond is initially considered to be creditworthy, and the chance that its financial health will decline increases as time passes. Unless the rating agency has made a mistake, there is very little chance an investment grade firm will default in the first year. By the fourth year, however, its business environment and financial condition may have changed.

For the CCC/C bonds, the probability of default per year is a decreasing function of time for the first few years. This bond (on average) is in the opposite position to an investment grade bond because the firm is considered to have problems at the time it receives its rating. If it survives the first year, there is a good chance its financial condition has improved. Surviving the first two years is even better news for its expected future financial condition. The probability of default, therefore, declines with the passage of time.

These tables indicate that S&P has done a reasonably good job. The numbers in each column of Table 4.3 increase as we move down from the AAA rating category to the CCC/C rating category. This means that the probability of a firm defaulting by year *n* increases as the initial credit rating decreases for all *n*. The corresponding tables for Moody's and Fitch show a similar pattern.

Hazard Rates

We showed that the conditional probability a firm initially rated B will default in the fifth year is 0.0280 (or 2.80%). Here, the future period we are considering is one year. Suppose we make the future period we are considering very short (i.e., of length Δt).

We can ask: "What is the probability of a firm defaulting between times t and $t + \Delta t$ conditional on no default before time t." This is $h\Delta t$, where h is the rate at which defaults are happening at time t. The variable h is referred to as the *hazard rate* (or the *default intensity*) at time t and is commonly used by analysts. It can be a function of time; in some models it is assumed to be stochastic (i.e., varying randomly through time).

Hazard rates are attractive because they can be used to calculate unconditional default probabilities. Suppose that $\bar{h}$ is the average hazard rate between time zero and time t. The unconditional default probability between time zero and time t is

$$1 - exp(-\bar{h}t)$$

The probability of survival (or survival rate) to time t is

$$exp(-\bar{h}t)$$

The unconditional probability of default between times t_1 and t_2 is

$$exp(-\bar{h}_1 t_1) - exp(-\bar{h}_2 t_2)$$

where $\bar{h}_1$ and $\bar{h}_2$ are the average hazard rates between today and times t_1 and t_2 (respectively).

As an example, suppose that the hazard rate is constant at 1% per year. The probability of a default by the end of the third year is

$$1 - exp(-0.01 \times 3) = 0.0296$$

or 2.96%. The unconditional probability of a default occurring during the fourth year is

$$exp(-0.01 \times 3) - exp(-0.01 \times 4) = 0.00966$$

or 0.966%. The conditional probability of defaulting in the fourth year, given that it has survived until the end of the third year, can be calculated by dividing the unconditional probability of a default occurring during the fourth year by the probability of surviving to the end of the third year. It is

$$0.00966/(1 - .0296) = 0.00995$$

or 0.995%.

From Table 4.3, we see the cumulative default probability over five years for a B-rated firm was 17.33% (or 0.1733). The average hazard rate over the five years is obtained by solving

$$1 - exp(-\bar{h} \times 5) = 0.1733$$

This gives

$$\bar{h} \times 5 = -ln(1 - 0.1733)$$

$\bar{h} = 0.0381$, or 3.81% per year.

4.3 RECOVERY RATES

A bankruptcy leads creditors to file claims against a firm. Sometimes the assets of the firm are liquidated so that the claimants can receive partial payments. Sometimes there is an agreed upon reorganization in which the claimants agree to their claims being adjusted in some way.

Lenders are ultimately interested in estimating the amount of money they could lose when they lend money to a firm. The probability of default (given by data such as that in Tables 4.3 to 4.5) is one important input to estimating this amount. The other is the recovery rate. If a firm does default, how much can the lender expect to recover?

The recovery rate for a bond is usually defined as the value of the bond shortly after default and it is expressed as a percentage of its face value. The loss given default, which provides the same information, is the percentage recovery rate subtracted from 100.

The expected loss from a loan over a certain period is

$$\text{Probability of Default} \times \text{Loss Given Default}$$

or

$$\text{Probability of Default} \times (1 - \text{Recovery Rate})$$

Bonds vary according to their seniority (i.e., the extent to which the bond holder will rank above other claimants in the event of default) and to the extent to which there is security (i.e., whether collateral has been posted). Statistics produced by rating agencies show the average recovery rate varies from around 25% for junior bonds (i.e., bonds that are subordinate to others) to around 50% for senior secured bonds.

An important point is that recovery rates are negatively correlated with default rates. During a recessionary period, default rates on bonds are high and recovery rates are low. During times when the economy is doing well, default rates on bonds are low and recovery rates are high. This is because the value of the defaulting firm's assets tends to be low (high) during bad (good) economic conditions. This negative correlation creates additional difficulties for bond portfolio managers: High default rates are doubly bad because they are accompanied by low recovery rates.

4.4 CREDIT SPREADS AND RISK PREMIUMS

Suppose you are determining what the interest rate should be on a five-year bond issued by a BBB-rated firm. You could look at Table 4.3 and see that there is a 1.64% chance the bond will default in the next five years. You then estimate the recovery rate on the bond as 40%. Your expected loss on the bond is then 1.64% × 0.6 or 0.984% of the bond's principal. This is the expected loss over five years. It therefore seems reasonable to charge extra interest over the risk-free rate of 0.984%/5, or about 20 basis points, to compensate for credit risk.

The extra interest is known as a *credit spread*. In practice, it is likely to be much more than 20 basis points. Perhaps five times as much. (The 20 basis point estimate is the actuarial compensation necessary for credit risk.) However, bond holders need a *risk premium* in addition to the actuarial compensation. The reason they need a risk premium is that bonds do not default independently of each other. During some years, the default rate is high and during other years it is low. This creates systematic risk, in other words, risk bond portfolio managers cannot diversify away. A risk premium of 80 to 100 basis points on BBB bonds is not unusual. It is high in relation to the actuarial risk premium of 20 basis points, but it is low when compared to the expected risk premium of 5% to 10% (500 to 1,000 basis points) on equities.

Another possible reason for a relatively high credit spread is liquidity. Bonds are relatively illiquid instruments, and bond traders may require compensation for difficulties they will experience when they try to sell bonds.

4.5 THE RATING PROCESS

Rating agencies rate publicly traded bonds and money market instruments. An instrument is typically first rated when it is issued; the rating is also reviewed periodically (usually at least every 12 months). The rating, based on a mixture of analysis and judgement, is only provided when there is information of sufficient quality for the rating agency to form an opinion. To quote S&P, "[t]he analysis generally includes historical and projected financial information, industry and/or economic data, peer comparisons, and details on planned financings" (S&P Global). In addition, the analysis is based on qualitative factors, such as the institutional or governance framework. A meeting with management is also commonly undertaken.

The fee for the rating is paid by the firm being rated. This is generally a few basis points applied to the notional amount of the bond.[4] If a firm chooses not to pay the fee, the rating agency may not issue a rating. This fee arrangement has been criticized because the rating agency is paid by the issuer even though the product it provides is used by the purchaser. It would make sense for the purchaser to pay, but this is organizationally difficult because of what is termed the "free rider" problem; once one investor has bought the rating, others may obtain it from that investor for free.

It is sometimes argued that when the issuer is paying, the rating agency will be inclined to assign a rating that the issuer thinks it deserves. The counterargument to this is that the rating agency relies on its reputation and therefore will not hesitate to give a low rating when it is warranted.

Outlooks and Watchlists

In addition to the ratings themselves, rating agencies provide what are termed *outlooks*. These are indications of the most likely direction of the rating over the medium term. A positive outlook means that a rating may be raised, a negative outlook means that it may be lowered, and a stable outlook means it is not likely to change. A developing (or evolving) outlook means that while the rating may change in the medium term, the agency cannot (as of yet) determine the direction of this change.

Placing a rating on a *watchlist* indicates a relatively short-term change is anticipated (usually within three months). Watchlists can be positive (indicating a review for a possible upgrade) or negative (indicating a review for a possible downgrade).

Rating Stability

Rating stability is an important objective for rating agencies. One reason for this is that bond traders are major users of ratings. Often, they are subject to rules governing the credit ratings of the bonds they hold (e.g., some bond portfolio managers are only allowed to hold investment grade bonds). If ratings changed frequently, bond traders might have to trade more frequently (and incur high transaction costs) to maintain the required bond rating distributions in their portfolios.

Another reason for rating stability is that ratings are used in financial contracts and (in some countries) by financial regulators.

[4] For example, S&P has quoted its fee as 6.75 basis points.

Frequent changes in ratings would cause problems. For example, a bond rated A− might be acceptable as collateral in a contract, whereas one rated BBB+ is not. If a bond's rating switched frequently between these two rating categories, it would create difficulties in the administration of the underlying contract.

As a result of this need for stability, ratings only change when rating agencies believe there has been a long-term change in a firm's creditworthiness.

Through-the-Cycle versus Point-in-Time

Economies are cyclical and vary from periods of high growth to periods of lower growth (or even contraction). Furthermore, a firm's probability of default changes in tandem with economic conditions.

Rating agencies must therefore decide whether to rate firms "through-the-cycle" or at a "point-in-time." A through-the-cycle rating tries to capture the average creditworthiness of a firm over a period of several years and should not be unduly affected by ups and downs in overall economic conditions. By contrast, a point-in-time rating is designed to provide the best current estimate of future default probabilities. In theory, a through-the-cycle estimate will understate the probability of default during the down part of the economic cycle and overstate it during the up part of the cycle.

Consistent with their desire to produce stable ratings, rating agencies produce through-the-cycle estimates. It is therefore not always the case that ratings worsen when the economy is doing poorly (or improve when the economy is doing well). Sometimes, rating users adjust ratings produced by rating agencies so that they are converted from through-the-cycle to point-in-time. To do this, it is necessary to find a way to index the health of the economy and then apply this measure to the through-the-cycle ratings. The index is applied in such a way that ratings increase when the economy is doing well and decrease when it is doing poorly. These adjustments must be calibrated to empirical data on defaults at various times during the economic cycle.

Industry and Geographic Consistency

Rating agencies use the same scales to characterize default risk across different industries and different countries. An important question is whether ratings are consistent. For example, does a BBB+ rating for a firm in a certain industry in California mean the same as a BBB+ rating for a firm in different industry in Germany?

Moody's, S&P, and Fitch are based in the United States and much of the information they report is based on U.S. data. The rating history for firms outside the United States is shorter than for firms in the United States. As a result, it is sometimes difficult to determine whether ratings for non-U.S. firms are consistent with those of U.S. firms. S&P provides statistics like those in Table 4.3 separately for firms in the United States, Europe, and emerging markets. The U.S. data considers default experience over a 15-year period as in Table 4.3, the European data considers default experience for a seven-year period, and the emerging markets data considers default experience over just five years. Table 4.6 compares the five-year default percentages for the three groups. It shows that a rating for a European firm has historically been better than the same rating for a U.S. firm. This is particularly true for firms with investment grade ratings. Indeed, the probability of default for a European firm with an investment grade rating has been around one-third of that for a U.S. firm.

The statistics for emerging markets firms show a different pattern from those for European and U.S. firms. Those rated AAA, AA and A had a very low five-year default probability. However, firms rated BBB fared slightly worse than U.S. firms and much worse compared to European firms. Because rating agencies are continuously striving for geographic consistency, we do not necessarily expect these performance differences to be the same in the future as they have been in the past.

There is less available data on the consistency of ratings across industries. In the past, it has been true that banks with a given rating show higher default rates than non-financial corporations with the same rating. Also, there has been less agreement among different rating agencies for banks than for other firms. Again, it is worth stressing that rating agencies strive for consistency and that extrapolating from past data is dangerous. Differences observed in the past may not always be observed in the future.

Table 4.6 **Five-year Cumulative Default Probabilities for United States, European, and Emerging Markets Reported by S&P in 2016**

Initial Rating	U.S. Firms	European Firms	Firms in Emerging Markets
AAA	0.41	0.00	0.00
AA	0.43	0.20	0.00
A	0.69	0.26	0.04
BBB	1.92	0.56	2.24
BB	7.89	3.71	5.26
B	18.70	12.43	12.10
CCC/C	51.42	43.37	26.32
Investment Grade	1.12	0.34	1.45
Speculative Grade	16.47	10.14	9.35
All rated	7.47	2.62	5.87

Source: 2018 Annual Global Corporate Default And Rating Transition Study © 2019, reproduced with permission of S&P Global Market Intelligence LLC and Standard & Poor's Financial Services LLC.

4.6 ALTERNATIVE TO RATINGS

The ratings produced by credit rating agencies such as Moody's, S&P, and Fitch can be contrasted with the information provided by organizations such as KMV (which is now part of Moody's) and Kamakura. These organizations use models to estimate default probabilities and provide the output from these models to clients for a fee. These models can include factors such as:[5]

- The amount of debt in the firm's capital structure,
- The market value of the firm's equity, and
- The volatility of the firm's equity.

In the simplest version of the model, default can occur at just one future time. The default happens if the value of the assets falls below the face value of the debt repayment that is required at that time. If V is the value of the assets and D is the face value of the debt, the firm defaults when $V < D$. The value of the equity at the future time is

$$\max(V-D, 0)$$

This shows that the equity is a call option on the assets of the firm with a strike price equal to the face value of the debt. The firm defaults if the option is not exercised. The probability of this can be calculated from standard option pricing theory (which will be covered in later chapters).

KMV and Kamakura provide point-in-time estimates and do not have the stability objective of rating agencies. It can be argued the output from these models responds to changing

[5] The underlying model is due to R. Merton, "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," *Journal of Finance*, 29 (1974): 449–470.

circumstances much more quickly than the ratings provided by agencies. Equity prices, which are a key input to their models, are continually changing to reflect the latest information. Ratings, on the other hand, are only reviewed periodically.

4.7 INTERNAL RATINGS

Banks and other financial institutions develop their own internal rating systems based on their assessment of potential borrowers. They typically base their ratings on several factors (e.g., financial ratios, cash flow projections, and an assessment of the firm's management). In general, each factor is scored, and then a weighted average score is calculated to determine the overall final rating.

It is important for banks to develop their own internal rating procedures for several reasons. First, external ratings are not always available. Second, regulatory credit risk capital depends on probabilities of defaults (PDs). Finally, the accounting standard IFRS 9 (and its FASB counterpart) require banks to take default probabilities into account when loans are valued on the balance sheet.

Like external ratings, internal ratings can be either through-the-cycle or point-in-time. There is a tendency for them to be point-in-time, but through-the-cycle ratings may be more relevant for relatively long-term lending commitments. It can be argued that regulators should encourage banks to use through-the-cycle ratings (e.g., by insisting that through-the-cycle probabilities be used for the determination of regulatory capital).[6] This is because point-in-time ratings are pro-cyclical (i.e., they may accentuate economic cycles). During bad economic conditions, point-in-time probabilities of default increase and banks become less inclined to lend. This makes it difficult for firms to fund working capital and fixed assets; this in turn can cause economic conditions to worsen further. During good economic conditions, the reverse happens and economic conditions are helped by an easing of credit.

Banks must back-test their procedures for calculating internal ratings. This typically requires at least ten years of data and involves producing something equivalent to Table 4.3. If the default statistics show that firms with higher ratings have performed better than those with low ratings, then a bank can have some confidence in its rating methodology.

Some banks are currently trying to automate their lending decisions using machine learning. With this approach, an algorithm is given a great deal of historical data on firms and whether they have defaulted. This is used to come up with a rule for distinguishing between those firms that default and those that do not. Arguably, the first attempt to do something like this was proposed by Altman in 1968.[7] He developed what has become known as the Z-score. Using a statistical technique known as discriminant analysis, he looked at the following ratios:

X_1: Working capital to total assets,

X_2: Retained earnings to total assets,

X_3: Earnings before interest and taxes to total assets,

X_4: Market value of equity to book value of total liabilities, and

X_5: Sales to total assets.

For publicly traded manufacturing firms, the Z-score was:

$$Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 0.999X_5$$

A Z-score above 3 indicated that the firm was unlikely to default. As the Z-score was lowered, the probability of default increased to the point where a firm with a Z-score below 1.8 had a very high probability of defaulting.

Today's machine learning algorithms use far more than five input variables and far more data than that used by Altman. Furthermore, the discriminant function does not have to be linear.

4.8 RATING TRANSITIONS

In addition to the information presented in Table 4.3, rating agencies produce rating transition matrices. These are tables showing the probability of a bond issuer migrating from one rating category to another during a one-year period. Table 4.7 shows the one-year rating transitions produced by S&P in its 2018 study (based on data taken from 1981 to 2018).

According to Table 4.7, an issuer that has just been given an A rating has an 88.17% probability of being A-rated one year later. On the upside, it also has a 1.69% chance and a 0.03% chance of being upgraded to AA and AAA (respectively). On the downside, it has a 5.16% chance and a 0.06% chance of being downgraded to BBB and defaulting (respectively).

The NR column indicates the probability that a firm is no longer rated at the end of a year. For analysis, it is often necessary

[6] The probabilities of default used to determine regulatory capital are through-the-cycle. However, IFRS 9 (and its FASB counterpart) require point-in-time estimates of probability of default when loans are valued.

[7] See E. I. Altman, "Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy," Journal of Finance 23, 4 (September 1968): 589–609.

Table 4.7 Shown are One-Year Global Rating Transitions Compiled by S&P

D denotes default, and NR denotes that the issuer was not rated at the end of the year.

	AAA	AA	A	BBB	BB	B	CCC/C	D	NR
AAA	86.99	9.12	0.53	0.05	0.08	0.03	0.05	0.00	3.15
AA	0.50	87.06	7.85	0.49	0.05	0.06	0.02	0.02	3.94
A	0.03	1.69	88.17	5.16	0.29	0.12	0.02	0.06	4.48
BBB	0.01	0.09	3.42	86.04	3.62	0.46	0.11	0.17	6.10
BB	0.01	0.03	0.11	4.83	77.50	6.65	0.55	0.65	9.67
B	0.00	0.02	0.08	0.17	4.93	74.53	4.42	3.44	12.41
CCC/C	0.00	0.00	0.11	0.20	0.59	13.21	43.51	26.89	15.50

Source: 2018 Annual Global Corporate Default And Rating Transition Study © 2019, reproduced with permission of S&P Global Market Intelligence LLC and Standard & Poor's Financial Services LLC.

to proportionally allocate the NR number to the other rating categories.

The table shows that investment grade ratings are more stable than non-investment grade ratings. For example, a AAA-rated firm has an 86.99% chance of staying AAA; a BB rated firm only has a 77.50% chance of staying BB. Meanwhile, firms in the CCC/C category only have a 43.51% chance of keeping their ratings during the year.

If we assume rating changes in successive years are independent (i.e. what happens in one year is not influenced by what happened in the previous year), we can calculate a transition matrix for *n* years using the transition matrix for one year (this involves matrix multiplication). The actual multi-year transition matrices (as reported by the rating agencies) are not quite the same as those calculated using this independence assumption. This is because of what is referred to as the *ratings momentum* phenomenon. If a firm has been downgraded in one year, it is more likely to be downgraded the next year. If a firm has been upgraded one year, it is more likely to be upgraded the next year.

Rating transition matrices are calculated for internal as well as external ratings. One test of ratings is whether rating transitions remain roughly the same from one year to the next. Research conducted several years ago suggests there are some differences across sectors, especially for investment grade issuers.[8] Transition matrices also seem to depend on the economic cycle. Specifically, downgrades increase significantly during recessions (this is despite the fact the ratings are designed to be through-the-cycle).

[8] See for example Nickell, Perraudin, and Varotto, "Stability of Rating Transitions," Journal of Banking and Finance (2000), 24, 1-2: 203–227.

4.9 ARE CREDIT RATING CHANGES ANTICIPATED?

An interesting question is whether ratings have information content. It is possible that when a rating is moved down, new information is being provided to the market so that both the stock and bond prices decline while credit default swaps spreads increase. It is also possible that the market has anticipated the information, making the rating agency a follower rather than a leader.

Researchers who have investigated this question have produced mixed results. Most agree that the stock and bond markets' reactions to downgrades are significant. This is particularly true when the downgrade is from investment grade to non-investment grade. However, the market's reaction to upgrades is much less pronounced. Part of the reason why downgrades impact prices while upgrades do not is that downgrades (particularly those from investment to non-investment grade) affect the willingness of investors to hold bonds. Also, these firms may have entered into contracts involving rating triggers, and a downgrade may have negative implications for them.

Hull, Predescu, and White (2004) look at the impact of rating changes on credit default swap spreads.[9] They examined outlooks and watchlists as well as rating changes. They found watchlist reviews for a downgrade contain significant information, but downgrades and negative outlooks do not. Positive rating events were much less significant. Generally, credit default swap changes seem to anticipate rating changes.

[9] See J. C. 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 (Nov 2004): 2789–2811.

Indeed, they found credit spread changes provide helpful information in estimating the probability of negative credit rating changes. They found that 42.6% of downgrades, 39.8% of reviews for downgrades, and 50.9% of negative outlooks come from the top quartile of credit default swap spread changes.

4.10 THE RATING OF STRUCTURED PRODUCTS

During the run-up to the 2007–2008 crisis, rating agencies became much more involved in the rating of structured products created from portfolios of subprime mortgages. A key difference between rating structured products and the traditional business of the rating agencies (i.e., rating regular bonds and money market instruments) is that the rating of a structured product depends almost entirely on a model. The rating agencies were quite open about the models they used; S&P and Fitch based their ratings on the probability that the structured product would give a loss, while Moody's based its ratings on expected loss as a percent of the principal. Unfortunately, the inputs to their models (particularly the correlations between the defaults on different mortgages) proved to be too optimistic, and their ratings of structured products created from other structured products proved to be questionable.

Once the creators of structured products understood the models used by rating agencies, they found that they could design the structured products in a way that would achieve the ratings they desired. In fact, they would present plans for structured products to rating agencies and get advanced rulings on ratings. Products that did not receive the desired rating were adjusted until they did.

Rating agencies found the work they were doing on structured products to be very profitable and they (perhaps) were not as independent as they should have been. As is well known, many of the structured products created from mortgages defaulted during the 2007–2008 crisis period and the reputation of rating agencies suffered as a result. As already mentioned, rating agencies are now subject to more oversight than before and (at least in the United States) are no longer used by bank supervisors to determine regulatory capital.

SUMMARY

The three largest rating agencies are Moody's, S&P, and Fitch. They have been in existence for around one hundred years and have developed a great deal of expertise in rating bonds and money market instruments. However, they have been less successful in rating structured products.

The performance of rating agencies can be assessed by examining how firms that were assigned a certain rating in a certain year subsequently performed. The tables reproduced in this chapter are indicative of a good performance by rating agencies in rating corporate bonds. There is some evidence the performance of bonds with a given rating varies across the United States, Europe, and emerging markets. There is also some evidence for variation across sectors. However, rating agencies are continually striving for consistency and these variations may not persist in the future.

Banks and other financial institutions have developed their own internal credit ratings that are similar to those of rating agencies. While the ratings produced by rating agencies are through-the-cycle, internal credit ratings can be either point-in-time or through-the-cycle. Point-in-time ratings are produced using models by firms such as KMV and Kamakura.

Analysts are interested in the probability of a firm moving from one rating category to another in a certain period. Firms with poor credit ratings are more likely to be subject to a rating change than firms with good credit ratings.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

QUESTIONS

Short Concept Questions

4.1 What ratings by Moody's, S&P, and Fitch for bonds are investment grade?

4.2 What ratings by Moody's, S&P and Fitch for bonds are speculative grade?

4.3 What is the Moody's rating equivalent to S&P's BBB+?

4.4 Is Fitch's rating scale for bond's most like that of Moody's or S&P?

4.5 How does a money market instrument differ from a bond?

4.6 What are the prime and non-prime ratings used by S&P for money market instruments?

4.7 How is the hazard rate at a particular time defined?

4.8 What is the typical relationship between the average recovery rate and the default rate?

4.9 What is the difference between through-the-cycle and point-in-time ratings? Which do rating agencies supply? What do KMV and Kamakura supply?

4.10 Why do bond portfolio holders need to earn a risk premium?

Practice Questions

4.11 What do rating transitions measure? How stable have rating transitions been through time?

4.12 What is the evidence on whether rating changes are anticipated?

4.13 In terms of default probabilities, how consistent have credit ratings been between different geographies? How would expectations for a European BBB-rated firm differ from those of a BBB-rated firm in the United States or in the emerging markets?

4.14 If the hazard rate is 1.5% per year for the first three years and 2.5% per year for the next three years, what is the probability of default during the first two years? What is the average hazard rate for the first five years? What is the probability of default between years two and five?

4.15 Using the data in Table 4.3, calculate the average hazard rate for a bond initially rated BB over ten years.

4.16 What is the difference between the way structured products are rated and the way corporate bonds are rated?

4.17 Using the data in Table 4.3, what is the credit spread on a ten-year AA-rated bond necessary to compensate for the actuarial probability of default. Why is the credit spread we see in the market greater than this?

4.18 Explain why equity can be regarded as an option on the assets of a firm.

4.19 What is the ratings momentum effect? Why is it relevant to determining rating transition matrices over periods longer than one year?

4.20 How might machine learning be used to automate lending decisions at some banks?

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

ANSWERS

Short Concept Questions

4.1 Investment grade bond ratings are BBB− and above (for S&P and Fitch) or Baa3 and above (for Moody's).

4.2 Speculative grade bond ratings are of BB+ or below (for S&P and Fitch) or Ba1 and below (for Moody's).

4.3 Baa1

4.4 Fitch's rating scale is like that of S&P.

4.5 A money market instrument lasts a year or less and provides all its return at the end of its life. A bond provides coupons (usually every six months).

4.6 Prime ratings include A-1+, A-1, A-2, and A-3. Non-prime ratings include B, C, and D.

4.7 The hazard rate h at a specific time is defined so that $h\Delta t$ is the probability of default between times t and $t + \Delta t$ conditional on no earlier default.

4.8 They are negatively correlated. When default rates are high, recovery rates are low. When default rates are low, recovery rates are high.

4.9 Point-in-time ratings reflect current economic conditions, whereas through-the-cycle ratings are longer term and are supposed to be independent of the economic cycle. Rating agencies provide through-the-cycle ratings. KMV and Kamakura provide point-in-time ratings.

4.10 They require a risk premium to compensate for systematic risk. During some years, default rates on bonds are high, reducing their returns. During other years, they fare much better. The low liquidity of bonds may also be a factor.

Solved Problems

4.11 Rating transitions measure the probability that a company will move from one rating to another during a certain period. Even though ratings are through-the-cycle, downgrades tend to be higher during recessionary times.

4.12 The evidence is unclear. Upgrades do not seem to provide significant news to the market. Downgrades can have an impact on stock prices and bond prices, but this may be because ratings are used for many different purposes (e.g., to define the mix of bonds that can be held by a portfolio manager or what bonds are acceptable as collateral).

4.13 European companies (particularly those with investment grade ratings) have performed better than U.S. companies. Emerging Markets companies with AA and A ratings have performed relatively well.

4.14 The probability of default during the first two years is $1\text{-}\exp(-0.015 \times 2) = 0.02955$. The average hazard rate during the first five years is $(1.5 \times 3 + 2.5 \times 2)/5 = 1.9\%$. The probability of default during the first five years is $1\text{-}\exp(-0.019 \times 5) = 0.09063$. The probability of default between years two and five is $0.09063 - 0.02955 = 0.06107$.

4.15 The cumulative default probability is 12.22%. The average hazard rate h is given by $-\ln(1 - 0.1222)/10 = 0.0130$ or 1.30%.

4.16 Structured products are rated using a model, whereas corporate bonds are rated using a mixture of judgement and analysis.

4.17 The cumulative ten-year default probability is 0.73%. This suggests that a credit spread as small as 7.3 basis points is sufficient to compensate for the actuarial probability of default. In practice, bond traders need compensation for systematic risk and possibly liquidity risk.

4.18 If the value of the assets falls below the face value of the debt that is due to be repaid at a future time, equity is in theory worth zero. Otherwise, it is worth the excess of the value of the assets over the face value of the debt. The equity can therefore be regarded as a call option on assets with a strike price equal to the face value of the debt.

4.19 The ratings momentum effect is that if a bond is downgraded (upgraded) in one year, it is more likely to be downgraded (upgraded) the next year. This means that rating transition matrices in successive years are not statistically independent.

4.20 If a bank has data describing millions of potential borrowers, and data on whether they default, it can use machine learning algorithms to predict defaults and make lending decisions.

5

Country Risk: Determinants, Measures, and Implications

Learning Objectives

After completing this reading, you should be able to:

- Explain how a country's economic growth rates, political risk, legal risk, and economic structure relate to its risk exposure.
- Evaluate composite measures of risk that incorporate multiple components of country risk.
- Compare instances of sovereign default in both foreign currency debt and local currency debt and explain common causes of sovereign defaults.
- Describe the consequences of sovereign default.
- Describe factors that influence the level of sovereign default risk; explain and assess how rating agencies measure sovereign default risks.
- Describe the characteristics of sovereign credit spreads and sovereign credit default swaps (CDS) and compare the use of sovereign spreads to credit ratings.

Many large firms have business interests all over the world. Though these firms may be incorporated in a single country, it is important for them to assess the risks associated with the foreign countries they operate in. These risks are collectively referred to as *country risk*.

There are numerous components of country risk. One such component is political risk. For example, a change in government might lead to a less welcoming regulatory environment for foreign companies. In some extreme cases, there is the risk that a firm's assets in a foreign country will be seized after a political coup. Other risks may be concerned with the inadequacies of a country's legal system, the level of corruption, and the potential for violence.

Individuals and corporations can obtain diversification benefits by investing outside their domestic markets. One way an individual can do so is by investing in those firms in his or her country with interests throughout the world. However, a more direct way is through an investment in one or more funds specializing in certain countries or regions. However, the managers of these funds need to evaluate country risk when making their investments. An important question is: What return is required when investing in country X? Investments in country X may be riskier than investments in the domestic market, necessitating an extra expected return (risk premium).

Governments finance themselves by issuing bonds and money market instruments as well as by borrowing from large international banks. When lending to foreign governments, it is important for lenders to consider country risk as part of their credit default risk framework. To that end, the credit rating agencies we considered in the previous chapter provide ratings assessing the probability of a country defaulting. These agencies must consider whether the taxes a government collects will be sufficient to meet all its obligations, including those related to servicing outstanding debt.

We start this chapter by taking a big picture view of country risk. We look at several services that provide scores for countries by considering corruption, violence, legal risk, political risk, and overall risk. We then move on to explore the history of sovereign defaults as well as consider the performance and stability of sovereign credit ratings.

5.1 EVALUATION OF RISK

Investors can add foreign investments to their portfolios more easily than ever before. For example, an investor in the United States can buy an emerging market exchange-traded fund (ETF) from Vanguard or Blackrock with a relatively low expense ratio. The fund manager will then invest in the equities of countries such as the Philippines, Egypt, India, China, South Korea, Hungary, Pakistan, Malaysia, Russia, Colombia, Poland, Indonesia, Peru, Turkey, Brazil, Thailand, Romania, Chile, South Africa, Croatia, and Mexico.[1]

Corporations have similar choices for the countries they can invest in. Many developing markets have economies that are growing faster than those of developed markets. However, this fast growth may be accompanied by higher economic risks and less stable political climates. Furthermore, there are often links between political and economic risks. If economic growth in a country slows, for example, dissatisfaction may lead to political turmoil and even lower future growth.

GDP Growth Rates

The growth of a country's economy is measured by its Gross Domestic Product (GDP). GDP is the total value of goods and services produced by all the people and firms in a country. Economists usually look at the growth rate in GDP after allowing for the impact of inflation (this is called the real GDP growth rate). For example, if the growth rate measured in domestic currency is 3% per year and inflation in the domestic currency is 2% per year, the real growth rate is only 1% (= 3% − 2%) per year.

An important consideration in assessing country risk is how a country will react to economic cycles. During economic downturns, for example, developing countries often see larger declines in GDP than their developed counterparts. This is because developing economies tend to rely more heavily on commodities. This means that they get squeezed by lower prices and demand during global recessions. For example, in 2009 the United States experienced its worst recession since the 1930s and saw a 2.8% decline in real GDP. At the same time, Mexico's real GDP shrunk by 4.7%. This was because Mexico's economy relied heavily on exports to its northern neighbor. On the other hand, some developing countries survived the 2007–2009 recession quite well. For example, China's real GDP growth rate was above 6% in 2009.

Table 5.1 gives the real GDP growth rates of several developing and developed countries in 2019 (compiled from the IMF World Economic Outlook database). The table shows developing countries such as Vietnam, India, and China did well in 2019 and achieved real GDP growth rates of more

[1] Bloomberg ranks countries according to their attractiveness to foreign investors using more than a dozen criteria. The rankings vary markedly from year to year. The countries given here are listed in order of their Bloomberg ranking in 2019. See https://www.bloomberg.com/news/articles/2019-05-30/emerging-market-scorecard-favors-growth-stars-as-trade-war-bites

Table 5.1 Positive and Negative Real GDP Growth Rates in 2019

Country	Real GDP Growth Rate (%)	Country	Real GDP Growth Rate (%)
Vietnam	6.50	Barbados	−0.10
China	6.14	Namibia	−0.18
India	6.12	Angola	−0.27
Philippines	5.72	Ecuador	−0.48
Kenya	5.60	Puerto Rico	−1.15
Egypt	5.52	Macao SAR	−1.32
Indonesia	5.04	Sudan	−2.62
Poland	4.03	Argentina	−3.06
Pakistan	3.29	Equatorial Guinea	−4.64
Israel	3.13	Nicaragua	−5.04
New Zealand	2.51	Zimbabwe	−7.08
United States	2.35	Islamic Republic of Iran	−9.46
Spain	2.18	Libya	−19.06
Greece	1.98	Venezuela	−35.00
Korea	1.95		
Norway	1.93		
Portugal	1.91		
Netherlands	1.77		
Australia	1.71		
Denmark	1.70		
Austria	1.61		
Canada	1.55		
France	1.25		
United Kingdom	1.24		
Belgium	1.21		
Russia	1.08		
Sweden	0.94		
Japan	0.89		
Brazil	0.88		
Switzerland	0.76		
South Africa	0.66		
Singapore	0.55		
Germany	0.54		
Mexico	0.40		
Turkey	0.25		
Lebanon	0.20		
Italy	0.01		

Source: IMF World Economic Outlook, www.imf.org

than 6%. On the other hand, Libya and Venezuela had the worst GDP growth rates. These numbers illustrate some of the risks associated with developing countries. While Venezuela is rich in oil, it has been held back by political turmoil and internal strife.

Political Risk

Political risk is the risk that changes in governments, decisions made by governments, or the way governments operate will significantly affect the profitability of a business or an investment.

Assessing political risk is never easy. Sometimes a change in leadership can transform the political landscape of a country and increase multiple forms of risk. As a result, investors tend to value government stability.

Whether democracies or authoritarian governments create more political risk is debatable. Authoritarian governments often do not change as frequently as democratically elected governments and can therefore pursue the same policies for a longer period. However, when there is a coup and one dictator takes over from another, there can be a sharp discontinuity in policy. In the case of democracies, governments may change relatively often, but the impact of the change is not usually as dramatic as it is with authoritarian governments.

There have been studies examining whether authoritarian governments lead to faster GDP growth than democratic governments. The results are mixed, with some studies arguing that countries with democratic governments grow faster and others arguing that countries with authoritarian governments grow faster. Some authors argue economic growth creates the desire for a democratic government, and so the causality may be the opposite of that assumed in other studies.

One aspect of political risk is corruption. When a business invests in a foreign country, it will almost certainly have to deal with the government bureaucracy. In some countries, as a practical matter, it may be impossible to do business without bribing public officials. This is a problem because bribery is illegal in many developed countries. For example, the Foreign Corrupt Practices Act (FCPA) is a statute in the United States prohibiting the bribing of foreign officials by Americans.

Bribes are an implicit tax on income that reduce profitability and returns for businesses operating in a country (and for investors in those businesses). The amount of money spent on bribes is typically uncertain (adding to the risk), and a firm may suffer (both financially and reputationally) if it is prosecuted in its home country.

Transparency International is an organization that uses surveys of experts living and working in different countries to compile a corruption index. The lower the index, the more corrupt the country is perceived to be. Table 5.2 shows the top 20 and bottom 20 countries in 2019.

Table 5.2 **Corruption Indices Produced for the Top 20 and Bottom 20 Countries in 2019**

Country	Corruption Index	Country	Corruption Index
Denmark	87	Nicaragua	22
New Zealand	87	Cambodia	20
Finland	86	Chad	20
Singapore	85	Iraq	20
Sweden	85	Burundi	19
Switzerland	85	Congo	19
Norway	84	Turkmenistan	19
Netherlands	82	Rep of Congo	18
Germany	80	Guinea Bissau	18
Luxembourg	80	Haiti	18
Iceland	78	Libya	18
Australia	77	Korea, North	17
Austria	77	Afghanistan	16
Canada	77	Equatorial Guinea	16
United Kingdom	77	Sudan	16
Hong Kong	76	Venezuela	16
Belgium	75	Yemen	15
Estonia	74	Syria	13
Ireland	74	South Sudan	12
Japan	73	Somalia	9

Source: Transparency International, www.transparency.org

Table 5.3 Global Peace Indices, 2019

Country	Peace Score	Country	Peace Score
Iceland	1.072	Venezuela	2.671
New Zealand	1.221	Mali	2.710
Portugal	1.274	Israel	2.735
Austria	1.291	Lebanon	2.800
Denmark	1.316	Nigeria	2.898
Canada	1.327	Sudan	2.921
Singapore	1.347	Ukraine	2.950
Slovenia	1.355	North Korea	2.995
Japan	1.369	Turkey	3.015
Switzerland	1.375	Pakistan	3.072
Czechia	1.383	Russia	3.093
Ireland	1.390	Republic of Congo	3.218
Australia	1.419	Libya	3.285
Finland	1.488	Central African Republic	3.296
Bhutan	1.506	Afghanistan	3.300
Malaysia	1.529	Yemen	3.369
Netherlands	1.530	Syria	3.412
Belgium	1.533	South Sudan	3.526
Sweden	1.533	Iraq	3.573
Norway	1.536	Somalia	3.574

Courtesy of the Global Peace Index, Produced by The Institute for Economics and Peace. Used by Permission.

Another aspect of political risk is violence. Violence makes it difficult for businesses to operate, leads to higher insurance costs, and may lead to a difficult or totally unsatisfactory work environment for employees. An index measuring violence is the Global Peace Index. Table 5.3 shows scores for the best and worst countries in 2019. In this case, low scores are better than high ones.

Many countries rank similarly in Tables 5.2 and 5.3. As a result, doing business in countries such as Denmark, New Zealand, Switzerland, and Canada should lead to few of the corruption and violence problems we have mentioned. By contrast, many African and Middle Eastern countries are not as easy to operate in.

Another source of political risk is nationalization or expropriation. This is a particularly significant problem for firms working with natural resources. For example, a firm may own a mine or have the rights to drill for oil in a country. However, if the firm is profitable and the host country's economy is faltering, it can be a popular policy for a government to seize the foreign firm's assets while paying very little compensation. Of course, this may have the effect of discouraging future investment in the country, and thus hurting its economy in the long term.

Legal Risk

Legal risk is the risk of losses due to inadequacies or biases in a country's legal system. A legal system that is trusted and perceived to be fair helps a country to attract foreign investment. Because business activities inevitably generate legal disputes, firms do not want to invest in a country where the legal system is biased, subject to government interference, and/or slow to the point of ineffectiveness. In this context, it is interesting to note that disputes between Russian oligarchs tend to be heard in British (rather than Russian) courts.[2]

[2] The highest profile one is Roman Abramovich vs. Boris Berezovsky, but there have been several others.

Table 5.4 International Property Rights Index for Sample Countries in 2019

Country	Overall Index	Legal and Political Index	Physical Property Index	Intellectual Property Index
Argentina	5.09	4.55	5.41	5.30
Australia	8.36	8.15	8.28	8.66
Brazil	5.56	4.35	6.08	6.26
Canada	8.26	8.40	8.27	8.12
China	6.03	4.93	7.15	6.02
Germany	7.85	7.66	7.60	8.29
Ghana	5.75	5.31	5.90	6.03
India	5.82	4.88	6.61	5.97
Indonesia	5.41	4.74	7.07	4.14
Kenya	5.15	4.06	6.49	4.89
Mexico	5.23	3.59	6.15	5.94
Russia	4.99	3.66	5.91	5.39
United Kingdom	8.04	7.79	7.87	8.47
United States	8.20	7.48	8.34	8.78

Source: The Property Rights Association, www.internationalpropertyrightsindex.org

Property rights and contract enforcement are important aspects of a legal system. For example, if an investor is to buy shares issued by a firm domiciled in a foreign country, it is important for him or her to know that the country has fair and well-thought-out rules concerning the rights of shareholders and firm governance. Specifically, the country should have a legal system where a firm and its management can be sued in the event of insider trading, actions that hurt shareholders, or attempts to deceive the market about the firm's financial health.

The International Property Rights Index is published by the Property Rights Alliance to help individuals and firms understand the risks they are taking when they invest abroad. A country's total index is made up of a Legal and Political Index (which is subdivided into Rule of Law, Political Stability, and Control of Corruption), a Physical Property Index (which is subdivided into Property Rights, Registering Property, and Ease of Access to Loans), and an Intellectual Property Index (subdivided into Intellectual Property Protection, Patent Protection, and Copyright Protection). The overall index in 2019 varied from Yemen (2.67) to Finland (8.71). Table 5.4 gives examples of the indices that were compiled in 2019 for a few countries from different regions of the world. (Larger numbers indicate a better legal system.)

The Economy

It is important to understand the economic risks associated with investing in a foreign country. GDP per capita and the real GDP growth rate tell part of the story, but it is also important to assess the country's competitive advantages and its level of economic diversification. For example, some countries are highly dependent on a single commodity. If the price of that commodity declines, the country and the value of its currency will suffer. Many African and Latin American countries fall into this category.

Countries, like firms, can develop competitive advantages. To quote Michael Porter, "[a] country's competitive advantage depends on the capacity of its industry to innovate and upgrade."[3] Based on several years of research, Porter argues there are four key determinants of a country's competitive advantage.

1. *Factor Conditions*: The nation's position in factors of production (such as skilled labor or infrastructure) necessary to compete in an industry.

[3] See M. E. Porter, "The Competitive Advantage of Nations," *Harvard Business Review*, March–April, 1990. https://hbr.org/1990/03/the-competitive-advantage-of-nations

2. *Demand Conditions*: The nature of home-market demand for the industry's product or service.
3. *Related and Supporting Industries*: The presence (or absence) of supplier industries and other related industries that are internationally competitive.
4. *Firm Strategy, Structure, and Rivalry*: The conditions governing how firms are created, organized, and managed, as well as the nature of domestic rivalry.

Examples of countries developing competitive advantages are Hong Kong, Singapore, South Korea, and Taiwan (known collectively as the four Asian tigers). They maintained real GDP growth rates greater than 7% per year between the 1960s and 1990s. Hong Kong and Singapore have since become world-leading international financial centers, while South Korea and Taiwan are world leaders in manufacturing and information technology.

An important consideration when investing in a country is the extent to which its economy is diversified. Large countries such as Brazil, India, and China can broaden their economic bases without too much difficulty. Some small countries, however, rely on a small number of goods or services. This makes them very susceptible to changes in the demand for the goods and services they produce. In theory, countries should be able to hedge their risks by entering into long-term contracts with other countries for the sale of these goods (combined perhaps with long term contracts for the goods they need to import), but in practice this does not seem to happen to any great extent.

Another consideration here is a country may face a trade-off between short- and long-term growth. In the short-term, growth may be maximized by focusing on the extraction and export of a commodity. But for sustainable long-term growth, it might be preferable to develop other industries. In this respect, it is interesting to note that Saudi Arabia has an ambitious plan to restructure the Kingdom's oil-dependent economy by privatizing state assets and diversifying its focus.[4]

5.2 TOTAL RISK

We have presented a few different components of country risk. However, it is natural to ask whether there exists a reliable composite risk measure (the equivalent of a VaR or expected shortfall for countries). There are some services that attempt to do this. One is Political Risk Services (PRS), which uses 22 measures of political, financial, and economic risk to calculate its index.

[4] See for example http://www.arabnews.com/node/1157931/business-economy

Individual firms can customize the PRS forecasting model to their own projects or exposures by adjusting the weighting attached to each of the variables, adding or subtracting variables, or otherwise tailoring the model to emphasize specific potential sources of risk.

Media outlets, such as Euromoney and The Economist, also provide country risk scores. Euromoney bases its scores on a survey of 400 economists, whereas The Economist develops country risk scores internally based on currency risk, sovereign debt risk, and banking risk. The World Bank also provides country risk data measuring corruption, government effectiveness, political stability, regulatory quality, the rule of law, and accountability.

It is difficult to compare these services because they use different scoring methods and consider different attributes of country risk. It is also important to keep in mind that not every dimension of country risk is necessarily relevant to every individual or corporation. There is also the question of scaling. (If one country has a score twice as high as another, in what sense is it twice as risky?)

In many ways, the rankings of countries are more important than their numerical scores. Taken together, these services provide a useful narrative of risk in various countries. In some cases, the narrative accompanying a score is more important than either the score or the ranking itself.

5.3 SOVEREIGN CREDIT RISK

One measure of a country's risk is the risk it will default on its debt. There are two types of sovereign debt: the type issued in a foreign currency (such as the USD) and the type issued in the country's own currency. We will consider each in turn.

Foreign Currency Defaults

Debt issued in a foreign currency is attractive to global banks and other international lenders. The risk for the issuing country, however, is that it cannot repay the debt by simply printing more money. To illustrate, the United States government can repay the debt it has issued in USD by printing more USD. This is referred to as increasing the money supply, and it may lead to inflation. A country such as Argentina, however, cannot do this when it has borrowed USD.

There have been many defaults on sovereign debt over the last 200 years. Table 5.5 shows foreign currency sovereign defaults that happened between 2010 and 2018 (as reported by Moody's). Most of the defaults involved the exchange of old bonds for new bonds with some net present value loss to lenders. Greece was the biggest borrower to default during the

Table 5.5 **Sovereign Defaults 2010–2018 Where at Least Some Foreign Currency Was Involved (For Greece and Cyprus the Euro Is Counted as a Foreign Currency)**

Country	Date	Debt Amount (Billions of USD)
Jamaica	Feb, 2010	7.9
Greece	Mar, 2012	261.5
Belize	Sep, 2012	0.5
Greece	Dec, 2012	42.1
Jamaica	Feb, 2013	9.1
Cyprus	Jul, 2013	1.3
Ukraine	Oct, 2015	13.3
Mozambique	Apr, 2016	0.7
Mozambique	Feb, 2017	0.7
Belize	Mar, 2017	0.5
Rep of Congo	Jul, 2017	0.3
Venezuela	Nov, 2017	31.1
Barbados	Jun, 2018	3.4

Source: Moody's.

2010–2018 period. Moody's estimated the first 2012 Greek default resulted in investor losses of more than 70%, whereas the second default resulted in losses of more than 60%.[5]

Defaults are caused by a combination of financial, economic, and political issues that were largely unforeseen at the times the loans were originally made. Many of the largest defaulters have been South American countries, and some have defaulted multiple times in the last 200 years.[6] A default may make it impossible for a government to finance itself for a certain period. Over the longer term, however, debt markets have proved to be remarkably forgiving.

Local Currency Defaults

Some countries have defaulted on debt issued in their own currency as well as on debt denominated in foreign currency. Two examples of this include the defaults of Brazil and Russia in 1990 and 1998 (respectively).

Research from Moody's indicates that countries are increasingly defaulting on both types of debt simultaneously. Why would countries default on debt denominated in their own currencies when they could simply print more money? There are several reasons.

- In the decades prior to 1971, currencies had to be backed by gold reserves. The amount of these reserves therefore limited a country's ability to print more money.
- Greece and other members of the European Union use the euro as their domestic currency. They do not, however, have the right to print euros (this is the responsibility of the European Central Bank). This means that Greece could not have solved its debt problems in 2012 by printing money.
- Printing more money debases the currency and leads to inflation. If firms in a country have foreign currency debt, debasing the value of the local currency (in which they earn their profits) can make it very difficult for them to repay this debt as it would have negative consequences for the local economy.

On the other hand, printing money is likely to be attractive in the short term because a country's reputation and credit rating will not immediately suffer.

Rating agencies, as we will see, typically provide both local currency ratings and foreign currency ratings. The local currency rating for a country is an opinion about the possibility of a country defaulting on its local currency debt. The foreign currency rating is an opinion about the possibility of it defaulting on its foreign currency debt. The local currency rating is typically one or two notches higher than the foreign currency rating.[7]

Impact of a Default

When a firm defaults, its creditors usually have the right to force it to liquidate. While they may end up getting less than the face value of the debt, at least the situation is resolved. When a country defaults, however, it cannot be liquidated. Usually the old debt is replaced by new debt or is restructured in some other way (e.g., by lowering the principal, lowering the interest payments, or extending the life of the debt). In the past, a default on debt might have been followed by military action (such as what happened to Venezuela in the 1900s). However, this does not happen in the modern era.

The modern consequences of a default by a sovereign nation include:

- A loss of reputation along with an increased difficulty in raising funds for several years,
- A lack of investors willing to buy the debt and equity of corporations based in the country,
- An economic downturn, and
- Political instability as the population loses faith in its leaders.

[5] A default by Argentina in 2014 is sometimes included, but this was what is referred to as a technical default. The required funds were deposited into a trustee account, but legal proceedings in the United States prevented the disbursement of the funds.

[6] Argentina defaulted in 1830, 1890, 1915, 1930, 1982, and 2001. Brazil defaulted in 1826, 1898, 1914, 1931, and 1983. Paraguay defaulted in 1827, 1874, 1892, 1920, 1932, and 1986.

[7] As explained in Chapter 4, rating agencies use modifiers to provide a finer rating scale. One notch is the difference between two ratings on this finer rating scale. S&P's rating of BB+ is one notch above BB. Moody's Ba3 rating is one notch above B1.

Researchers have also found that a default negatively affects GDP growth, negatively affects the country's credit rating for many years, can hurt exports, and can make the defaulting country's banking systems more fragile.

In short, defaulting on debt is not something countries should take lightly. It can seriously impede their economic development and growth. Often, the International Monetary Fund becomes involved in restructuring the debt and imposing strict austerity conditions on the defaulting country. For example, it can insist government budget deficits are reduced through spending cuts, tax increases, or a combination of both.

Typically, defaults happen only after a country has experienced very difficult economic conditions or political upheavals. For example, a large default by Argentina in 2001 was a result of an economic depression starting in the third quarter of 1998. This caused widespread unemployment, riots, and the fall of a government.

5.4 SOVEREIGN CREDIT RATINGS

Rating agencies look at several factors when rating countries.[8] It is obviously important to consider the amount of debt a country already has. In the case of a firm, the ratio of the book value of debt to the book value of equity is a well-known leverage ratio. Because there is no equity measure for countries, however, rating agencies instead look to the ratio of government debt to GDP.

Table 5.6 shows statistics on the debt-to-GDP ratio for large developed and developing countries in 2018. Japan appears to be an outlier in these statistics, but there is a reason for this. The Japanese government, to a much greater extent than other governments, holds assets. These assets include cash, securities, and real estate holdings. The Japanese government also holds some of its own bonds.[9] When the figures are adjusted for those asset holdings not earmarked for other purposes (e.g., pension payments), the debt-to-GDP ratio becomes much more reasonable. Indeed, depending on exactly how the adjustments are made, the debt-to-GDP ratio becomes somewhere in the 40% to 90% range.

It can be relevant to look at how debt ratios have changed over time. In the United States, the debt-to-GDP ratio was in the 30% to 40% range between 1966 and 1984. It was in the 40% to 60% range for most of the period between 1985 and 2005. By 2018, the ratio had risen to over 100%.

[8] See for example https://www.spratings.com/documents/20184/774196/How+We+Rate+Sovereigns.pdf for how the S&P rating agency rates countries.

[9] These bonds are referred to as JGBs, Japanese Government Bonds. It is not clear why a government would buy its own bonds.

Table 5.6 Debt as Percent of GDP in 2018

Country	Government Debt (% of GDP)
Japan	237.54
Greece	174.15
United States	106.70
France	99.20
Brazil	90.36
United Kingdom	85.67
India	69.04
Germany	56.93
China	55.36
Russia	13.79

Source: International Monetary Fund, www.imf.org

It is worth noting debt issued by a government is not a country's total indebtedness. In the United States, for example, states such as California, and municipalities such as New York City also borrow money. It is also useful to distinguish between debt held internally (by a country's own citizens, banks, or other corporations) and that held externally (by foreign investors).

There are several other factors that are considered when determining a rating.

- *Social Security Commitments*: Governments make commitments to their citizens to pay pensions and provide health care. As the size of these commitments increases, a government has less free cash to service debt.
- *The Tax Base*: A rating agency must assess the size and reliability of the tax base. Countries with diversified economies will tend to have a more stable tax base than those that depend on one or two industries.
- *Political Risk*: It is sometimes argued autocracies are more likely to default than democracies. Given that the alternative to default is printing money, the decision-making process at the country's central bank may be important. Thus, it is important to know the extent to which the central bank is independent of the government.
- *Implicit Guarantees*: Countries in the Eurozone may well be helped by rich member countries (e.g., Germany and France) when they get into financial difficulties. However, there is no explicit guarantee that this help will always be given.

Table 5.7 compares the default rate experienced by sovereign countries with the default rate experienced by corporations with the same rating for periods ending in 2018 (the data for corporations was considered in Chapter 4). Countries rated

Table 5.7 Comparison of Sovereign Foreign Currency and Sovereign Debt Denominated in the Local Currency

	Ten-Year Cumulative Default Rate		
Rating	**Corporations**	**Sovereign Foreign Debt**	**Sovereign Local Debt**
AAA	0.70	0.00	0.00
AA	0.73	0.00	0.10
A	1.28	5.20	5.42
BBB	3.44	4.76	3.94
BB	12.22	11.63	6.79
B	24.21	24.81	9.07
CCC/C	50.44	67.60	42.84
Investment Grade	1.96	1.91	2.29
Speculative Grade	20.62	19.15	9.35
All Rated	8.88	7.91	4.60

Source: 2018 Annual Global Corporate Default And Rating Transition Study © 2019, reproduced with permission of S&P Global Market Intelligence LLC and Standard & Poor's Financial Services LLC.

AAA and AA have performed very well over the following ten-year period, whereas countries with an A rating have not performed as well. Sovereign foreign debt with BBB, BB, and B ratings has performed about the same as corporate debt with those ratings, but sovereign local debt has performed better than corporate debt. A CCC to C rating for sovereign foreign debt has a much higher ten-year default probability than the same rating for corporate debt, but local debt with that rating has a lower ten-year default probability.

As mentioned earlier, a country's local debt is typically rated slightly higher than its foreign debt. According to the data, debt denominated in local currency has performed better than debt denominated in foreign currency in most categories. The ten-year default rate on the former is about half of that on the latter.

Table 4.7 of Chapter 4 shows one-year transition rates for corporations. In Table 5.8, we show a table of rating transitions for the foreign debt of countries. In Table 5.9, we do the same for local debt. The tables show the rating transitions of countries are more stable than those of corporations except for the lowest rating category. The probability that corporations rated AAA, AA, A, BBB, BB, and B will keep that rating for one year are 86.99%, 87.06%, 88.17%, 86.04%, 77.50%, and 74.53% (respectively). Tables 5.8 and 5.9 show the corresponding numbers are higher for both sovereign local debt and sovereign foreign debt.

Table 5.8 One-Year Rating Transitions for Debt Denominated in a Foreign Currency from 1975–2018. NR Indicates a Transition to the Not-Rated Category

	AAA	**AA**	**A**	**BBB**	**BB**	**B**	**CCC/CC**	**Default**	**NR**
AAA	96.6	3.3	0.0	0.0	0.1	0.0	0.0	0.0	0.0
AA	2.5	93.3	3.0	0.3	0.3	0.0	0.0	0.0	0.5
A	0.0	3.7	90.7	5.0	0.4	0.0	0.0	0.0	0.2
BBB	0.0	0.0	5.3	89.3	4.7	0.5	0.2	0.0	0.0
BB	0.0	0.0	0.0	6.3	86.5	6.0	0.6	0.5	0.1
B	0.0	0.0	0.0	0.0	5.2	87.7	3.0	2.8	1.2
CCC/CC	0.0	0.0	0.0	0.0	0.0	29.1	29.3	41.6	0.0

Source: 2018 Annual Global Corporate Default And Rating Transition Study © 2019, reproduced with permission of S&P Global Market Intelligence LLC and Standard & Poor's Financial Services LLC.

Table 5.9 **One-Year Rating Transitions for Debt Denominated in the Local Currency from 1993–2018. NR Indicates a Transition to the Not-Rated Category**

	AAA	AA	A	BBB	BB	B	CCC/CC	Default	NR
AAA	95.8	4.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AA	1.5	91.5	5.5	0.9	0.0	0.0	0.0	0.0	0.6
A	0.0	2.6	90.6	6.1	0.4	0.2	0.0	0.0	0.2
BBB	0.0	0.0	5.3	87.1	6.8	0.6	0.1	0.0	0.0
BB	0.0	0.0	0.0	5.4	84.7	7.9	1.0	0.7	0.2
B	0.0	0.0	0.0	0.0	5.7	88.5	2.6	1.6	1.6
CCC/CC	0.0	0.0	0.0	0.0	0.0	46.8	41.5	11.6	0.0

Source: *2018 Annual Global Corporate Default And Rating Transition Study* © 2019, reproduced with permission of S&P Global Market Intelligence LLC and Standard & Poor's Financial Services LLC.

5.5 CREDIT SPREADS

The credit spread for sovereign debt in a specific currency is the excess interest paid over the risk-free rate in that currency. There is a strong correlation between credit spreads and ratings, but credit spreads can provide extra information on the ability of a country to repay its debt. One reason for this is that credit spreads are more granular than ratings (i.e., there is a range of credit spreads associated with a credit rating). One country with a given rating might have a lower credit spread (and therefore be perceived as less risky) than another country with the same rating.

Credit spreads also have the advantage of being able to adjust more quickly to new information than ratings. However, they are also more volatile. Recall from Chapter 4 that rating stability is one of the objectives of rating agencies. This is true for country bond ratings as well as for corporate bond ratings. As mentioned in Chapter 4, a rating is changed only when there is reason to believe the long-term financial health of the firm or country has changed.

One source of credit spread data is the credit default swap market. Credit default swaps are traded on countries as well as corporations. They are like insurance contracts in that they provide a payoff to the holder if the country defaults within a certain period (usually five years). Roughly speaking, the payoff is designed to put the bondholder in the same position as he or she would have been if the bond had not defaulted. Unlike an insurance contract, a CDS can be used by speculators as well as hedgers. For example, an entity does not have to have exposure to Brazilian debt to buy protection against a Brazilian default. If Brazil does not default, the speculator pays a premium and receives no payoff. If Brazil does default, the speculator receives the corresponding payoff even if he or she does not have any real exposure.

There has been some controversy about the actions of speculators in the sovereign credit default swap market. Speculators were blamed by some for driving up Greek CDS spreads in 2010. It was claimed this also drove up the credit spread for bonds issued by Greece, making the country's financial problems more severe.

To see how this could happen, suppose the euro risk-free rate is 3% and the yield on Greek debt is 10%. If speculators have driven up the cost of protection on the debt to 9%, bond holders will consider a 10% yield to be too low because the close-to-risk-free return to someone who buys the debt and hedges with a CDS would be only 1% (= 10% − 9%). This is less than the 3% they could earn elsewhere.

The European Union became concerned about the potential for speculators to influence the bond market in the way we have just described. As a result, it introduced legislation banning the purchase of uncovered sovereign CDS contracts (i.e., CDS contracts where the purchaser does not own bonds issued by the country and has no exposure to a default by the country). However, it is important to note most researchers argue Greek bond yields were not affected by CDS spreads.

The bonds issued by a country may, in practice, be a better source of credit spread data than the CDS market. In a CDS, there is always the possibility the seller of protection will default (these concerns rose during the 2007–2008 credit crisis). The CDS market is also subject to illiquidity problems as well as clustering (i.e., where the CDS spreads of groups of countries move together in a way that does not necessarily reflect default risk).

SUMMARY

Country risk is important to firms considering expansion opportunities involving a foreign country. It is also important to fund managers investing in foreign markets, banks making loans to foreign countries, and those considering buying bonds issued by foreign countries.

There are many components to the evaluation of country risk. Less developed countries encounter many problems as they try to grow their economies. Political upheavals, reductions in the market price of key exports, and civil wars are all liable to adversely affect the foreign firms investing in the country.

There are several services that score a country according to its risks and, in some cases, provide an in-depth analysis. The risks considered include economic risk, political risk, legal risk, the level of violence, and so on. Sometimes the risks are combined into a single measure.

Rating agencies such as Moody's, S&P, and Fitch provide country ratings for debt issued in foreign currency and debt issued in the local currency. The default experience for sovereign debt with a certain rating is sometimes better and sometimes worse than that for corporate debt with the same rating. For most rating categories, local currency debt has performed better than foreign currency debt. The ratings for both local and foreign currency debt tend to be more stable than those for corporate debt.

Political upheavals, reductions in the market price of key exports, and civil wars are all liable to lead to a default on debt. Some countries, such as Argentina, Brazil, and Paraguay, have defaulted many times on their debt. A default typically leads to existing debt being exchanged for new debt, or the existing debt being restructured in some way (e.g., with a lower interest rate, a lower principal, or a longer time for repayments). Typically, a default makes it difficult for a country to raise further debt for a period of time and is likely to hurt both its reputation and its economy.

The credit spread for a country's debt is the excess of the interest rate charged over the risk-free rate for the currency in which the debt is denominated. The credit spread varies from country to country and tends to reflect the country's credit rating. However, credit spreads are more granular and more up-to-date measures of country risk than credit ratings.

The following questions are intended to help candidates understand the material. They are <u>not</u> actual FRM exam questions.

QUESTIONS

Short Concept Questions

5.1 What is the difference between the GDP growth rate and the real GDP growth rate?

5.2 Give three countries that rank in the top 20 both in terms of having the lowest levels of corruption *and* the lowest levels of violence.

5.3 Give three countries that rank in the bottom 20 for both corruption *and* violence.

5.4 How does diversification decrease a country's economic risk?

5.5 What are the two types of sovereign debt rated by ratings agencies?

5.6 Name three countries that have defaulted on their debt since 2010.

5.7 How does having a gold standard make it difficult for a country to service its debt?

5.8 Are sovereign credit ratings more or less stable than corporate credit ratings?

5.9 Is the ten-year default rate for AAA- and AA-rated corporate debt more or less than that for sovereign debt?

5.10 Which adjusts more quickly to new information: credit spreads or credit ratings?

Practice Questions

5.11 Did developed or developing countries have higher GDP growth rates in 2019?

5.12 What are the main categories of country risk for a business thinking of expanding into a foreign country?

5.13 Do authoritarian governments lead to greater country risk than democratic governments? Why or why not?

5.14 What are the important aspects of a country's legal risk?

5.15 Is local currency debt rated more highly or less highly than foreign currency debt? Discuss the reasons for your answer.

5.16 Why might a country decide to default on its local currency debt rather than print more money? Assume the country is not on the gold standard and could print more money if it was inclined to do so.

5.17 What are some possible consequences for a country defaulting on its debt?

5.18 Why is Japan's debt-to-GDP ratio so high?

5.19 What factors are considered in determining the rating of a South American country's debt?

5.20 Explain the relationship between a country's CDS spread and the credit spread on its bonds.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

ANSWERS

Short Concept Questions

5.1 The real growth rate in GDP is the growth rate adjusted for inflation.

5.2 There are many such as Iceland, New Zealand, Austria, Denmark, Canada, Switzerland, Japan, Australia, Norway, Finland, Sweden, Belgium, and the Netherlands.

5.3 There are many such as North Korea, Somalia, South Sudan, Syria, Yemen, Sudan, Libya, Afghanistan, Iraq, and Congo.

5.4 It is less likely to be affected by the ups and downs in the prices of particular goods.

5.5 Rating agencies provide a rating for debt denominated in the local currency and for debt denominated in a foreign currency.

5.6 The countries that have defaulted are Jamaica, Greece, Belize, Cyprus, Ukraine, Mozambique, Congo, Venezuela, and Barbados.

5.7 The currency issued had to be backed by holdings of gold.

5.8 Sovereign ratings are more stable (except for the lowest rating).

5.9 The ten-year default experience for AAA- and AA-corporate debt is greater than that for sovereign debt with the same rating.

5.10 Credit spreads adjust more quickly.

Solved Problems

5.11 As indicated by Table 5.1, developing countries had the highest and the lowest GDP growth rates. Developed countries were in the middle.

5.12 The main categories of country risk are economic risk, political risk, legal risk, and the level of violence.

5.13 Authoritarian governments change less frequently than democratic governments, but when they do change there can be extreme discontinuities in policy.

5.14 Property rights, the enforcement of contracts, and the time it takes to resolve disputes.

5.15 Local currency debt is more highly rated. A country can print money to avoid a default on local currency debt. This makes a default on local currency debt less likely.

5.16 Printing more money will lead to inflation, which will be bad for the economy.

5.17 A debt default makes it difficult for a country to raise money for several years. Investors will be reluctant to buy the debt or equity of firms domiciled in the country. The default may lead to an economic downturn and political instability.

5.18 The Japanese government owns assets more than other governments.

5.19 Debt-to-GDP ratio, social security commitments, the tax base, and political risk.

5.20 The CDS spread should, in theory, be close to the credit spread on the country's bonds. This is because an investor expects to earn close to the risk-free rate by buying bonds and buying CDS protection on the bonds.

6

Measuring Credit Risk

Learning Objectives

After completing this reading, you should be able to:

- Explain the distinctions between economic capital and regulatory capital and describe how economic capital is derived.
- Describe the degree of dependence typically observed among the loan defaults in a bank's loan portfolio, and explain the implications for the portfolio's default rate.
- Define and calculate expected loss (EL).
- Define and explain unexpected loss (UL).
- Estimate the mean and standard deviation of credit losses assuming a binomial distribution.
- Describe the Gaussian copula model and its application.
- Describe and apply the Vasicek model to estimate default rate and credit risk capital for a bank.
- Describe the CreditMetrics model and explain how it is applied in estimating economic capital.
- Describe and apply Euler's theorem to determine the contribution of a loan to the overall risk of a portfolio.
- Explain why it is more difficult to calculate credit risk capital for derivatives than for loans.
- Describe challenges to quantifying credit risk.

In this chapter, we continue our discussion of credit risk by considering risk in the loan portfolios held by banks. Specifically, our focus will be on the amount of equity capital banks need to mitigate this risk. The capital must be enough to absorb the losses from loans in nearly all circumstances.

It is important to distinguish between economic capital and regulatory capital. Economic capital is a bank's own estimate of the capital it requires. Regulatory capital is the capital bank regulators (also known as bank supervisors) require a bank to keep. Global bank regulatory requirements are determined by the Basel Committee on Banking Supervision in Switzerland. The requirements are then implemented by bank supervisors in each member country.[1]

We consider three different models for quantifying credit risk in this chapter. The first is a model where the mean and standard deviation of the loss from a loan portfolio is determined from the properties of the individual loans. The second model, known as the Vasicek model, is used by bank regulators to estimate an extreme percentile of the loss distribution. These extreme percentile estimates determine bank regulatory capital requirements. The third model is known as CreditMetrics and is often used by banks themselves when estimating economic capital.

6.1 BACKGROUND

A bank must keep capital for the risks it takes. When losses are incurred, they come out of the equity capital. For example, if a bank has equity capital of USD 5 billion and incurs losses of USD 1.5 billion, its equity capital is reduced to USD 3.5 billion. If losses are so large that the equity capital becomes negative, the bank is insolvent and may be unable to remain in business.

How much equity capital does a bank need? This depends on the losses it might incur. If losses as high as USD 5 billion are virtually impossible, then USD 5 billion of equity capital will probably be viewed as sufficient. However, if there is even a 1% chance of such a loss, both the bank and its regulators are likely to feel that additional capital is needed.

Banks have debt capital as well as equity capital. The debt capital is almost always subordinate to deposits. This means that if the equity capital is wiped out by losses, the debt holders should (in theory) incur losses before depositors are affected.[2]

Equity capital is sometimes referred to as "going concern capital" since as it is positive, the bank is solvent and can therefore be characterized as a going concern. By contrast, debt capital is referred to as "gone concern capital" since it only becomes an important cushion for depositors when the bank is no longer a going concern (i.e., insolvent).

Banks are subject to many risks. Trading operations give rise to market risks and the need to quantify these risks (as discussed in Chapters 1 to 3). Operational risks (such as cyber risks and litigation risks) are important and will be discussed in the next chapter. In this chapter, we focus on credit risk.

Credit risk has traditionally been the most important risk taken by banks (since the core activity of banking is taking deposits and making loans). Each loan issued by a bank has some risk of default and therefore a risk of a credit loss. It is the possibility of this loss that constitutes credit risk.[3]

The Basel Committee

Prior to 1988, there was no global banking regulation. Instead, each country chose how it regulated its own banks. Most set minimum ratios for capital to total assets. However, the definition of capital and the definition of assets varied from country to country.

By the 1970s and 1980s, banks were competing globally while their transactions were becoming increasingly complicated (e.g., the rise of over-the-counter interest rate swaps in the early 1980s). These trends encouraged national bank regulators to pursue a global approach.[4]

In 1974, the central banks of the G10 countries formed the Basel Committee to harmonize global bank regulation. By 1988, the committee had agreed on a common approach for determining the required credit risk capital for the banks under their supervision. This regulation is now known as Basel I. In 1996,

[1] See https://www.bis.org/bcbs/ for Basel Committee publications

[2] The governments of most countries have deposit insurance schemes to protect depositors. For example, the Federal Deposit Insurance Corporation (FDIC) in the United States provides protection to depositors up to USD 250,000 if a bank defaults. The fact that debt is subordinate to deposits means that debt holders should take losses before the FDIC does.

[3] Banks also face credit risk in other aspects of their business including foreign exchange transactions; trading in financial futures, swaps, bonds, and equities; and in the extension of credit commitments and loan guarantees.

[4] The first interest rate swap was arranged by Salomon Brothers in 1981 between IBM and the World Bank. IBM borrowed in Swiss francs and German marks. The World Bank borrowed in USD. IBM agreed to service the World Bank's debt, and in return the World Bank agreed to service IBM's debt. The swap was motivated by the fact that the World Bank was limited in the direct borrowings it was allowed to do in Swiss francs and German marks.

they extended Basel I by requiring capital for market risk as well as credit risk. In 1999, Basel II was proposed. It introduced a capital requirement for operational risk and changed the capital requirements for credit risk. Between 1999 and 2019, we have seen the implementation of Basel II and further regulations, known as Basel II.5 and Basel III.

The Basel II credit risk regulations, which still underlie credit risk capital calculations, feature a standardized approach and an internal ratings-based (IRB) approach. The standardized approach involves use of credit ratings and other metrics. In this chapter, we will explain the IRB approach, which is based on the work of Vasicek (1987) and Gordy (2003).[5]

Economic Capital

Economic capital is a bank's own estimate of the capital it requires. Both regulatory capital and economic capital feature separate capital calculations for credit risk, market risk, and operational risk.[6] For regulatory capital, the results are added to give the total capital requirements. For economic capital, however, correlations between the risks are often considered.

In this chapter, we will explain the CreditMetrics model for calculating credit risk economic capital. One of the features of this model is that it considers losses from both credit rating downgrades and loan defaults.

Data on Defaults

If defaults among a bank's borrowers were independent of each other, we could expect the default rate to be similar each year. For example, if each of the 100,000 loans in a bank's portfolio has a 1% chance of defaulting, the bank would experience around 1,000 defaults each year. There would also be a less than 0.1% chance that yearly defaults would exceed 1,100. (This can be shown using the BINOM.DIST function in Excel. Entering BINOM.DIST(1100,100000,0.01,TRUE) returns 0.9992.)

In practice, loans do not default independently of each other. Rather, there are good years and bad years for defaults. This is illustrated in Table 6.1, which shows the annual percentage default rate on all rated companies between 1981 and 2018 (as estimated by Moody's). The default rate varies from a low of 0.14% in 1981 to a high of 4.19% in 2009.

Table 6.1 **Annual Global Percentage Default Rates for All Rated Companies, 1981–2018**

Year	Default Rate	Year	Default Rate	Year	Default Rate
1981	0.14	1994	0.63	2007	0.37
1982	1.19	1995	1.05	2008	1.80
1983	0.76	1996	0.51	2009	4.19
1984	0.91	1997	0.63	2010	1.21
1985	1.11	1998	1.28	2011	0.80
1986	1.72	1999	2.14	2012	1.14
1987	0.94	2000	2.48	2013	1.06
1988	1.38	2001	3.79	2014	0.69
1989	1.78	2002	3.60	2015	1.36
1990	2.73	2003	1.93	2016	2.08
1991	3.25	2004	0.78	2017	1.20
1992	1.49	2005	0.60	2018	1.03
1993	0.60	2006	0.48		

Source: 2018 Annual Global Corporate Default And Rating Transition Study © 2019, reproduced with permission of S&P Global Market Intelligence LLC and Standard & Poor's Financial Services LLC.

The observations in Table 6.1 have a mean of 1.443% and a standard deviation of 0.984%. A lognormal distribution can be used to provide an approximate fit to the data. (A lognormal distribution is the distribution of a variable whose natural logarithm is normally distributed.) Converting the percentages in Table 6.1 to decimals and taking logarithms, we find the distributional mean for the logarithm of the default rate is −4.458, and the standard deviation is 0.699. The fitted lognormal distribution is shown in Figure 6.1.

An important reason why companies do not default independently of each other is the economy. The economy represents systematic risk that banks and bond holders cannot diversify away. Good economic conditions immediately before and during a year decrease the probability of default for all companies during the year; bad economic conditions immediately before and during a year increase the probability of default for all companies during the year. From Table 6.1, we can deduce that economic conditions were good immediately

[5] See O. Vasicek, "Probability of Loss on a Loan Portfolio," Working Paper, KMV, 1987 (Published in *Risk* in December 2002 under the title, "Loan Portfolio value," and M. B. Gordy, "A Risk-Factor Model Foundation for Ratings-Based Capital Ratios," *Journal of Financial Intermediation*, 12 (2003): 199–232.

[6] Sometimes economic capital is also calculated for strategic risk.

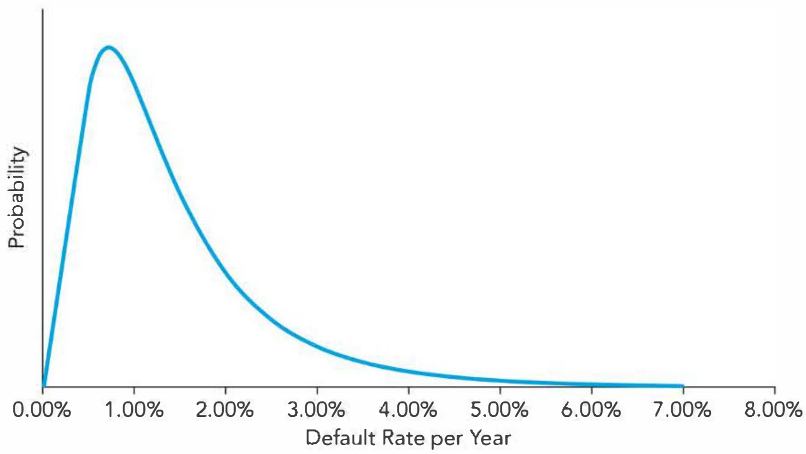


Figure 6.1 Lognormal distribution fitting the default data in Table 6.1.

before and during 1981. Meanwhile, 2009 was the tail end of a severe crisis, with the result being an extremely high rate of default.

As we will explain later, the Vasicek model relates the probability of default to the value of a factor. That factor can be considered to be a measure of the recent health of the economy.

What is known as credit contagion may also play a role in the year-to-year variations in Table 6.1. This is the process where a problem in one company (or one part of the economy) leads to problems in other companies (or other parts of the economy). To provide a simple example of credit contagion, suppose Company A buys goods from Company B, which in turn buys goods from Company C. If Company A goes bankrupt, Company B will suffer and, if the volume of transactions between the two is sufficient, may itself go bankrupt. This in turn could cause Company C to fail, even though it had no direct exposure to Company A.

The importance of credit contagion for non-financial companies is debatable. But there can be no question that bank regulators are very concerned about the potential for credit contagion in the banking system. This is referred to as systemic risk.[7] If Bank A fails, Bank B may take a huge loss because of the transactions (particularly the over-the-counter derivatives transactions) it has with Bank A. This could lead to Bank B defaulting. If Bank C has many outstanding transactions with both Bank A and Bank B, it might then experience financial difficulties as well.

The Model for Determining Capital

The model used to estimate both economic capital and regulatory capital for credit risk is illustrated in Figure 6.2. The figure shows the probability distribution of default losses during a one-year period. The expected loss (i.e., the mean of the distribution) and a high percentile ($X\%$) of the loss distribution are indicated.

Consider first the expected loss. This is something a bank allows for in the way it sets interest rates on its loans. As a simple example, assume a bank charges just enough on its loans to cover expected losses on the loans, the expenses involved in managing and administering the loans, and the cost of funding the loans. If the expected default rate on its loan portfolio is 1.5% and the recovery rate in the event of a default is 40%, the expected loss is 0.9% ($= 0.6 \times 1.5\%$). Suppose a bank needs a margin of 1.6% on its lending to cover its expenses. The total margin to cover its expected losses and expenses is therefore 2.5% ($= 1.6\% + 0.9\%$). If the bank's average funding cost is 1%, the average interest rate it charges on its loans needs to be 3.5% ($= 1.0\% + 2.5\%$).[8]

A challenge in managing credit risk is that actual losses do not equal expected losses every year (as illustrated by the data in Table 6.1). The 0.9% default loss on the loan portfolio in our example is simply an average. In some years, losses will be less than this; in other years, they will be greater. The bank's capital is a cushion that (as indicated in Figure 6.2) covers the unexpected loss (i.e., the actual loss in a given year above the expected loss). In earlier chapters, we learned about value-at-risk (VaR). Using this terminology, the required capital is the expected loss subtracted from the VaR with a one-year time horizon and an $X\%$ confidence level.

What should X be? The Basel Committee sets $X = 99.9\%$ for regulatory capital in the internal ratings-based approach. This is because it wants the capital for credit risk to cover losses 99.9% of the time. To put this another way, bank supervisors require capital to be sufficient for losses that (in theory) occur only once every thousand years.

When banks determine economic capital, they tend to be even more conservative. Consider a bank rated as AA. One of its key objectives will almost certainly be to maintain its AA credit rating.

[7] Systemic risk is not the same as systematic risk. Systemic risk is the risk of the collapse of the entire financial system. Systematic risk is the overall market risk created by ups and downs in the economy.

[8] To keep the example simple, we have not considered the required profit that the bank would, in practice, build into its calculations.

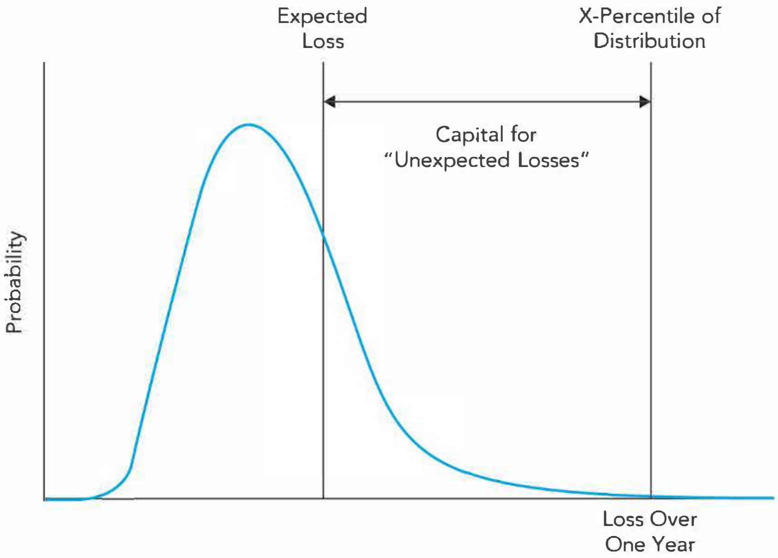


Figure 6.2 **Illustration of how capital is determined for a loan portfolio.**

We saw in Chapter 4 that an AA-rated corporation has a default probability of about 0.02% in one year. A bank can maintain an AA rating if it is able to convince rating agencies there is 99.98% probability that it will not default. One way of doing this is by setting X as high as 99.98% in Figure 6.2. (The capital is then theoretically sufficient to absorb losses in 4,999 out of 5,000 years!)

The rest of this chapter will be concerned with determining the X percentile of the loan loss distribution. Before we start, however, we can do some calculations based on the lognormal fit to the data in Table 6.1. The logarithm of the default rate has a mean of -4.458 and a standard deviation of 0.699. Assuming the loss distribution is lognormal, the 99.9 percentile of the distribution of the default rate logarithm is

$$-4.458 + (N^{-1}(0.999) \times 0.699) = -4.458 + (3.09 \times 0.699) = -2.298$$

where N^{-1} is the inverse cumulative normal distribution. The 99.9 percentile of the default rate distribution is therefore around 10.05% (= exp(−2.298)). Similarly, the 99.98 percentile of the distribution is around 13.76%. These would be the "worst case" default rates used in setting capital for a bank whose portfolio mirrors the set of all rated companies considered in Table 6.1.

When a loan defaults, there is usually some recovery made. Unfortunately (as mentioned in Chapter 4), there is a negative relationship between default rates and recovery rates.[9] When default rates are very high, we expect the recovery rates to be low. If we assume a recovery rate of 25%, the 99.9 and 99.98 percentiles of the loss distribution are 7.54% (= 0.75 × 10.05%) and 10.32% (= 0.75 × 13.76%), respectively.

The expected default rate for all rated companies (according to Table 6.1) is 1.443%. If we apply a recovery rate of 25% to this, we obtain an expected loss rate of 1.08% (= 0.75 × 1.443%).[10] The regulatory capital requirement would therefore be around 6.46% (= 7.54% − 1.08%) and the economic capital would be around 9.24% (= 10.32% − 1.08%).

These estimates assume that the bank's loan portfolio has the same risk as the average rated company. This may not necessarily be the case.

6.2 THE MEAN AND STANDARD DEVIATION OF CREDIT LOSSES

Suppose a bank has n loans, and define the following quantities:

- L_i: The amount borrowed in the ith loan (assumed constant throughout the year)
- p_i: The probability of default for the ith loan
- R_i: The recovery rate in the event of default by the ith loan (assumed known with certainty)
- ρ_{ij}: The correlation between losses on the ith and jth loan
- σ_i: The standard deviation of loss from the ith loan
- σ_P: The standard deviation of loss from the portfolio
- α: The standard deviation of portfolio loss as a fraction of the size of the portfolio

If the ith loan defaults, the loss is $L_i(1 - R_i)$. The probability distribution for the loss from the ith loan therefore consists of a probability p_i that there will be a loss of this amount, and a probability $1 - p_i$ that there will be no loss. This is a binomial distribution. We will now calculate the mean and standard deviation of the loss.

The mean loss is

$$p_i \times L_i(1 - R_i) + (1 - p_i) \times 0 = p_i L_i(1 - R_i)$$

[9] See also Nada Mora, "What determines creditor recovery rates," Federal Reserve Board of Kansas City, *Economic Review*, 79–109.

[10] It can be argued that the recovery rate for the loss in an average year should be higher than the recovery rate in a year that corresponds to an extreme percentile of the loss distribution.

The formula for the standard deviation of a random variable X is

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

where E denotes expected value. This leads to:

$$\sigma_i^2 = E(\text{Loss}^2) - [E(\text{Loss})]^2$$

As we have already pointed out, E(Loss) is $p_i L_i(1 - R_i)$. There is a probability p_i that the square of the loss will be $[L_i(1 - R_i)]^2$ and a probability $1 - p_i$ that that the square of the loss will be zero. This means that:

$$[E(Loss^2)] = p_i[L_i(1 - R_i)]^2$$

This means that:

$$\sigma_i^2 = p_i[L_i(1 - R_i)]^2 - [p_i L_i(1 - R_i)]^2 = (p_i - p_i^2)[L_i(1 - R_i)]^2$$

and

$$\sigma_i = \sqrt{p_i - p_i^2}[L_i(1 - R_i)]$$

We can calculate the standard deviation of the loss on the loan portfolio from the standard deviation of the losses on the individual loans. This is

$$\sigma_P^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{ij}\sigma_i\sigma_j \qquad (6.1)$$

The standard deviation as a percentage of the size of the portfolio is

$$\alpha = \frac{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{ij}\sigma_i\sigma_j}}{\sum_{i=1}^{n}L_i} \qquad (6.2)$$

We now simplify matters by assuming all loans have the same principal L, all recovery rates are the same and equal to R, all default probabilities are the same and equal to p, and:

$$\rho_{ij} = \begin{matrix} 1 & \text{when } i = j \\ \rho & \text{when } i \neq j \end{matrix}$$

where ρ is a constant. The standard deviation of the loss from loan i (σ_i) is then the same for all i. Denoting the common standard deviation by σ, we obtain

$$\sigma = \sqrt{p - p^2}[L(1 - R)]$$

Equation (6.1) gives[11]

$$\sigma_P^2 = n\sigma^2 + n(n - 1)\rho\sigma^2$$

[11] The summations in Equation (6.1) give rise to n^2 terms. Of these, there are n terms where $i = j$ and $n(n - 1)$ terms where $i \neq j$. Each of the terms where $i = j$ equals σ^2, and each of the terms where $i \neq j$ equals $\rho\sigma^2$.

The standard deviation of the loss from the loan portfolio as a percentage of its size (i.e., the standard deviation of the percentage loss) is

$$\alpha = \frac{\sigma_P}{nL} = \frac{\sigma\sqrt{1 + (n - 1)\rho}}{L\sqrt{n}}$$

To illustrate this formula, suppose a bank has a portfolio with 100,000 loans, and each loan is USD 1 million and has a 1% probability of default in a year. The recovery rate is 40%. In this case (using USD million as the unit of measurement) $n = 100{,}000$, $p = 0.01$, $R = 0.4$, and $L = 1$ so that:

$$\sigma = \sqrt{(0.01 - 0.0001)} \times 1 \times 0.6 = 0.0597$$

Table 6.2 shows values of α for different values of ρ. To illustrate these calculations, consider the case where $\rho = 0.1$. The parameter α is then:

$$\frac{0.0597\sqrt{1 + (99{,}999 \times 0.1)}}{\sqrt{100{,}000} \times 1} = 0.0189$$

Figure 6.3 displays the relationship between α and ρ. As expected, α increases as ρ increases. The rate of increase is greatest for small values of ρ.

To calculate bank capital using the model in Equation (6.1), we need to estimate correlations and then estimate the unexpected loss in the way indicated in Figure 6.2. Correlations can be estimated from a bank's historical data. Also, rating agencies estimate the correlations between the default rates of two companies as a function of their rating.

The calculation of unexpected loss requires an estimate of a high percentile of the loss distribution. There is no easy way to obtain this, however, and so the Monte Carlo simulation is often employed. On each simulation, random sampling determines whether each debtor defaults during a one-year period, and the results for many trials determine

Table 6.2 Standard Deviation of Percentage Loss, α, for Different Values of the Correlation Parameter ρ

ρ	α
0.00	0.0002
0.01	0.0060
0.02	0.0084
0.05	0.0134
0.1	0.0189
0.2	0.0267
0.5	0.0422
1.0	0.0597

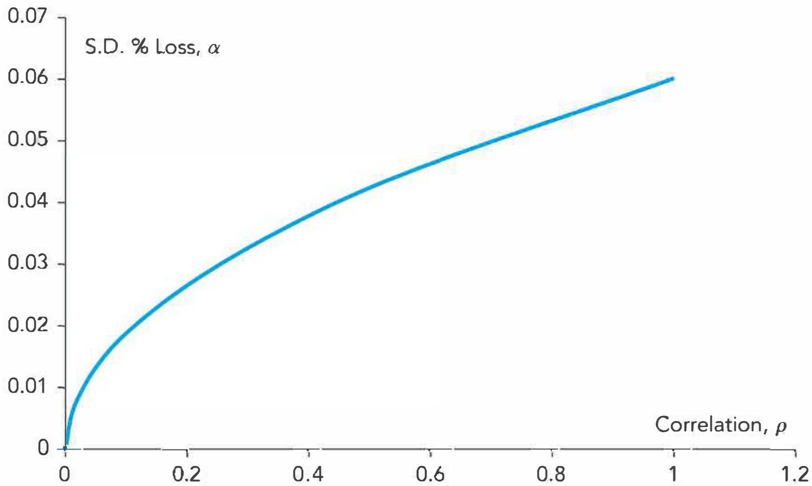


Figure 6.3 The impact of correlation, ρ, on the standard deviation of percentage loss, α, for a large portfolio where the probability of default (p) is 1%, and the recovery rate (R) is 40% for all loans.

a probability distribution for default losses during the year. The methodology is like the CreditMetrics approach (which will be explained in Section 6.5).

6.3 THE GAUSSIAN COPULA MODEL

Before moving on to describe the models used by bank supervisors and banks themselves, it will be useful to explain a tool known as the Gaussian copula model. This is because it is used in both bank and regulator models.

Suppose we know probability distributions for variables V_1 and V_2, and we also want to define the complete way in which they depend on each other (i.e., their joint probability distribution). We can assume the joint distribution is bivariate normal so long as both variables are normally distributed.[12] If their distributions are not normal, we can transform each distribution to a standard normal distribution by transforming their percentiles to the corresponding percentiles of a standard normal distribution. This is illustrated in Figure 6.4.

[12] The bivariate normal distribution is one way in which two normally distributed variables can be jointly distributed. Suppose that the variables have means μ_1 and μ_2 and standard deviations σ_1 and σ_2. Conditional on the first variable equaling V, the second variable is normal with:

$$\text{Mean} = \mu_2 + \rho\sigma_2\frac{V - \mu_1}{\sigma_1} \qquad \text{SD} = \sigma_2\sqrt{1 - \rho^2}$$

There are many other ways in which a joint distribution can be created from two normal distributions. For example, if the variables are V_1 and V_2, it could be the case that $V_2 = V_1$ for $-2 \leq V_1 \leq +2$, and $V_2 = -V_1$ otherwise. $V_1 + V_2$ is then not normal.

We suppose V_1 is transformed to U_1 and V_2 is transformed to U_2, where U_1 and U_2 are (by construction) normal distributions with mean zero and standard deviation one. The one-percentile point on the distribution for V_1 is transformed to the one-percentile point of the distribution for U_1 (i.e., -2.326), the five-percentile point on the distribution for V_1 is transformed to the five-percentile point on the distribution for U_1 (i.e., -1.645), and so on. Similarly, the one-percentile point on the distribution for V_2 is transformed to $U_2 = -2.326$, the five-percentile point on the distribution for V_2 is transformed to $U_2 = -1.645$, and so on.

We then assume the distributions of U_1 and U_2 are bivariate normal with a particular correlation.[13] This defines the joint distribution of U_1 and U_2. Because the transformations (V_1 to U_1 and V_2 to U_2) are both one-to-one transformations, it also unambiguously defines the joint distribution of the variables we are interested in (i.e., V_1 and V_2). The key point here is that we do not know how to define a reasonable joint distribution between variables V_1 and V_2 directly, so we transform them into variables that we can easily work with and define the joint distribution of the transformed variables.

A One-Factor Correlation Model

Now suppose we have many variables, V_i ($i = 1, 2, \ldots$). Each V_i can be mapped to a standard normal distribution U_i in the way we have described. We are then faced with the problem of defining the correlation between the U_i distributions. When there are many different distributions, this involves specifying a large number of different correlation parameters. Analysts often handle this problem by using a one-factor model.

A one-factor model assumes that:

$$U_i = a_iF + \sqrt{1 - a_i^2}Z_i \tag{6.3}$$

In this model, F is a factor common to all the U_i (this is the reason why it does not have a subscript), and Z_i is the component of U_i that is unrelated to the common factor F. The Z_i corresponding to the different U_i are uncorrelated with each other. The a_i are parameters with values between -1 and $+1$.

The variables F and Z_i have standard normal distributions (i.e., normal distributions with mean zero and standard deviation 1). The variable U_i is the sum of two independent normal distributions and is therefore normal. It has a mean of zero because it is the sum of two components, each of which has a mean of zero.

[13] See Footnote 12 for properties of the bivariate normal distribution. In applications of this model to determine credit risk capital, the correlation is specified by the Basel Committee.

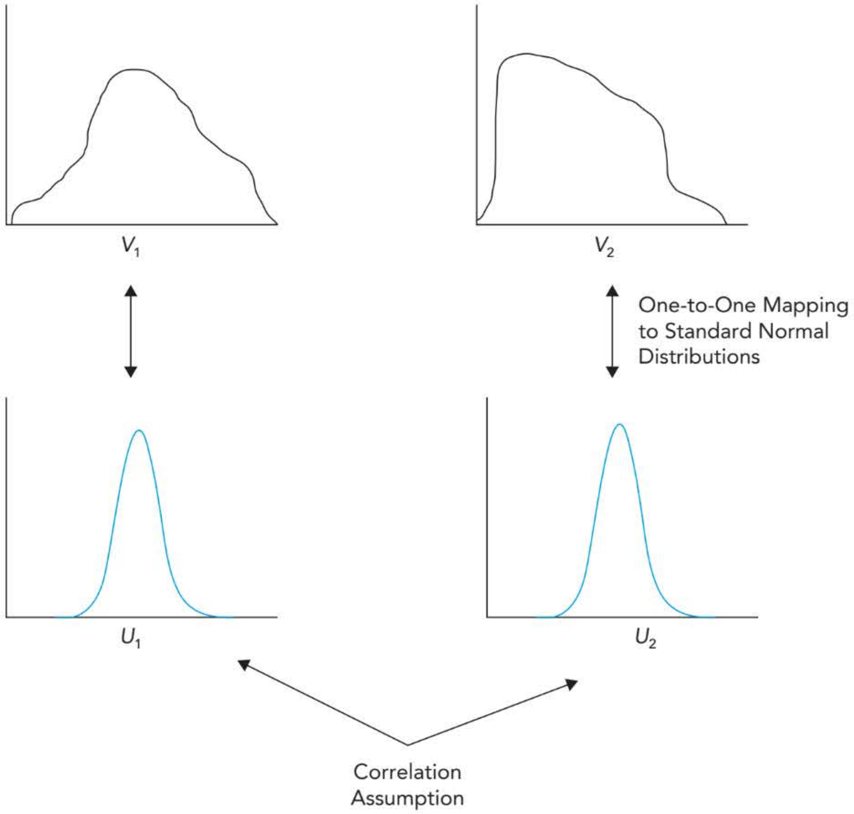


Figure 6.4 **Illustration of the Gaussian copula model.**

The variable U_i has a standard deviation of one. To see this, note the following.

- a_iF_i has a variance of a_i^2.
- $\sqrt{1 - a_i^2}Z_i$ has a variance of $1 - a_i^2$.
  F_i and Z_i are uncorrelated.
- The variance of the sum of two uncorrelated variables is the sum of their variances.

This explains the rather strange-looking coefficient of Z_i. Equation (6.3) takes a variable U_i with a standard normal distribution and defines it in terms of two other variables, each of which has a standard normal distribution. The first variable is the common factor F that affects all the U_i. The other is a variable Z_i that only affects U_i.

The coefficient of correlation between U_i and U_j arises only from their dependence on the common factor F and thus is a_ia_j. To show this, we first note that a standard formula in statistics gives the coefficient of correlation between U_i and U_j as:

$$\frac{E(U_iU_j) - E(U_i)E(U_j)}{SD(U_i)SD(U_j)}$$

Because $E(U_i)$ and $E(U_j)$ are zero, and $SD(U_i)$ and $SD(U_j)$ are both one, this reduces to:

$$E(U_iU_j)$$

or:

$$E\left[\left(a_iF + \sqrt{1 - a_i^2}Z_i\right)\left(a_jF + \sqrt{1 - a_j^2}Z_j\right)\right]$$

Because F is uncorrelated with all the Z_i, and Z_i is uncorrelated with Z_j:

$$E(FZ_i) = E(FZ_j) = E(Z_iZ_j) = 0$$

Thus, the coefficient of correlation between U_i and U_j becomes

$$E(a_ia_jF^2)$$

This is a_ia_j because $E(F^2) = 1$.

An example of a one-factor model is the capital asset pricing model (CAPM). In this model, the correlation between the returns from two stocks is assumed to arise entirely from their dependence on a common factor. In the case of CAPM, the common factor is the return from the market index. CAPM is an approximation. For example, there is presumably some correlation between the return from General Motors and the return from Ford that is unrelated to their correlation with the market index. However, CAPM makes the specification of correlations between the returns from different stocks easy to handle.

In the next section, the one-factor model described will be applied to default probabilities. The factor can be thought of as a variable related to the economy that affects default rates.

Before proceeding, it is worth emphasizing the Gaussian copula is not the only way of defining the joint distribution of the V_i. Many other, more complex copulas have been proposed. Tail correlation is the probability of extreme values of V_i and V_j happening at the same time. This varies according to the copula model chosen. The Gaussian copula has relatively little tail correlation and therefore requires relatively large correlation estimates to fit market data.[14]

[14] Prior to the 2007–2009 crisis, it was often used to model the default correlation between mortgages. It was later criticized because it underestimated the tail correlation.

6.4 THE VASICEK MODEL

We are now able to explain the Vasicek model, which is used by regulators to determine capital for loan portfolios. It uses the Gaussian copula model to define the correlation between defaults. The Vasicek model has the advantage over the model in Section 6.2 in that the unexpected loss can be determined analytically.

Assume the probability of default (PD) is the same for all companies in a large portfolio. The binary probability of the default distribution for company i for one year is mapped to a standard normal distribution U_i as described in the previous section.

Values in the extreme left tail of this standard normal distribution correspond to a default, whereas the rest of the distribution corresponds to no default. This is illustrated in Figure 6.5. Company i therefore defaults if:

$$U_i \leq N^{-1}(\text{PD})$$

where N^{-1} is the inverse cumulative normal distribution.

For example, if PD = 1%, company i defaults if:

$$U_i \leq N^{-1}(.01) = -2.326$$

Values for U_i between minus infinity and -2.326 correspond to default, while values between -2.326 and infinity correspond to no default.

The model in Equation (6.3) defines the correlation between defaults. To make the model manageable for a loan portfolio, the a_i are assumed to be the same for all i. Setting $a_i = a$, the model becomes

$$U_i = aF + \sqrt{1 - a^2}Z_i$$

and the correlation between each pair of U-distributions is a^2.

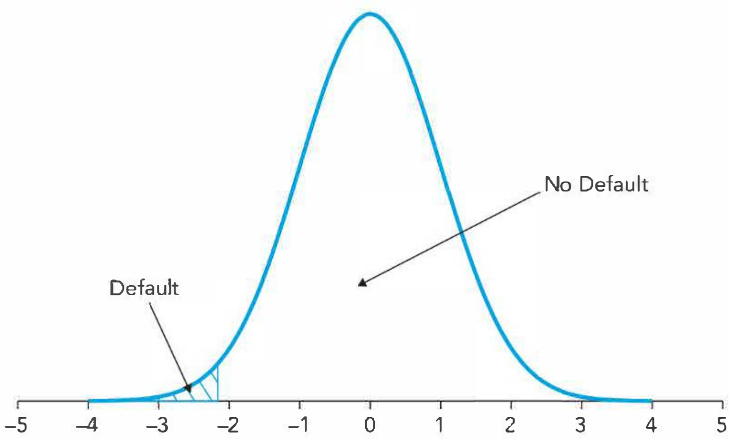


Figure 6.5 **Relation between the value of U_i and whether a default occurs.**

As indicated earlier, the factor F can be thought of as an index of the recent health of the economy. If F is high, the economy is doing well and all the U_i will tend to be high (making defaults unlikely). If F is low, however, all the U_i will tend to be low so that defaults are relatively likely. For each value of F, the distribution of each U_i has a mean of aF and a standard deviation of $\sqrt{1 - a^2}$. For a large portfolio, the default rate is the probability U_i is less than $N^{-1}(\text{PD})$. From the properties of normal distributions, we therefore get

$$\text{Default Rate as a function of } F = N\left(\frac{N^{-1}(PD) - aF}{\sqrt{1 - a^2}}\right) \quad (6.4)$$

Equation (6.4) is a powerful formula. For a large portfolio, it gives the relationship between the value of the economic factor F and the portfolio default rate.

The default rate, which we do not expect to be exceeded with a 99.9% probability, is given by a very low value of F, denoted by F^*. We require that the probability of the true value of F will be worse than F^* and will be 0.1%. Since F is normally distributed, $F^* = N^{-1}(0.001)$. We therefore obtain

$$\text{99.9 percentile for default rate} = N\left(\frac{N^{-1}(PD) - aN^{-1}(0.001)}{\sqrt{1 - a^2}}\right)$$

If we define ρ as the correlation between each pair of U_i distributions as $\rho = a^2$, this formula becomes[15]

$$\text{99.9 percentile for default rate} = N\left(\frac{N^{-1}(PD) - \sqrt{\rho}N^{-1}(0.001)}{\sqrt{1 - \rho}}\right)$$

Because $N(0.001) = -N(0.999)$, this can be written as:

$$\text{99.9 percentile for default rate} = N\left(\frac{N^{-1}(PD) + \sqrt{\rho}N^{-1}(0.999)}{\sqrt{1 - \rho}}\right) \quad (6.5)$$

For a loan portfolio where each loan has the same default probability, Equation (6.5) allows us to convert the average default rate (PD) into a portfolio default rate which should (in theory) happen only once every 1,000 years. Note that when $\rho = 0$, the formula gives a 99.9 percentile of the default rate equal to PD. This is what we would expect because (if companies default independently) there is a "law of large numbers" phenomenon at work, and the default rate in a large portfolio is always the same. Table 6.3 gives the relationship between PD and the 99.9 percentile of the default rate using Equation (6.5). As expected, the 99.9 percentile of the default rate increases with both the PD and the ρ.

[15] Given that the correlations are all the same they must be non-negative.

Table 6.3 Relation Between PD and the 99.9 Percentile of the Default Rate for a Large Portfolio of Loans

	PD = 0.1%	PD = 0.5%	PD = 1%	PD = 1.5%	PD = 2%
$\rho = 0.0$	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%

To illustrate the calculations in Table 6.3, consider the case where PD = 1.5% and $\rho = 0.4$. The 99.9 percentile for the portfolio default rate is

$$N\left(\frac{N^{-1}(0.015) + \sqrt{0.4}N^{-1}(0.999)}{\sqrt{1-0.4}}\right)$$
$$= N\left(\frac{-2.1701 + 0.6325 \times 3.0902}{\sqrt{0.6}}\right)$$
$$= N(-0.2784) = 39.0\%$$

Consider a portfolio of loans with the same PD, the same correlations ρ, the same loss given default (LGD), and the same principal.[16] The Basel II capital requirement for banks that use the IRB approach is:

$$(\text{WCDR} - \text{PD}) \times \text{LGD} \times \text{EAD} \tag{6.6}$$

where the WCDR (worst case default rate) is the 99.9 percentile of the default rate distribution given by Equation (6.5), LGD is the loss given default (equals one minus the recovery rate), and EAD is the total exposure at default (i.e., the sum of the principals of all the loans).

Equation (6.6) gives the unexpected loss with a 99.9% confidence level. This is because WCDR × LGD is the 99.9 percentile point of the loss rate distribution, and WCDR × LGD × EAD is the loss at this 99.9 percentile point. Similarly, PD × LGD × EAD is the expected loss.[17]

Work by Gordy (referenced earlier) shows that for a non-homogeneous loan portfolio, the one-factor model in Equation (6.6) can reasonably be extended so that the required capital is

$$\sum_i (\text{WCDR}_i - \text{PD}_i) \times \text{EAD}_i \times \text{LGD}_i \tag{6.7}$$

where WCDR_i, PD_i, EAD_i, and LGD_i are the values of WCDR, PD, EAD, and LGD (respectively) for loan i. This equation means the capital for each loan can be considered separately, and the results added together. For some types of loans, Equation (6.7) is modified to incorporate a maturity adjustment factor. This recognizes that for a loan lasting over one year, credit quality can deteriorate during the year without a default occurring.

The Basel II rules define the correlation ρ that banks must assume in different situations. Using the IRB approach, banks make PD estimates. The EAD and LGD estimates are (depending on the circumstances and the regulatory approvals) either determined in accordance with the Basel II rules or by using the bank's own models.[18]

A final point on the Vasicek model is that the correlation parameter it uses is not the same as the correlation parameter in the model in Section 6.2. Vasicek's correlation parameter is the correlation between the normal distributions into which the binary default-or-not distributions are transformed. The correlation parameter in Section 6.2 is between the binary probability distributions themselves.

6.5 CREDITMETRICS

CreditMetrics is the model banks often use to determine economic capital. Under this model, each borrower is assigned an external or internal credit rating. A one-year transition table (like that discussed in Chapter 4) is used to define how ratings change.

The bank's portfolio of loans is valued at the beginning of a one-year period. A Monte Carlo simulation is then carried out to model how ratings change during the year. In each simulation trial, the ratings of all borrowers at the end of the year are determined, and the portfolio is revalued. The credit loss is calculated as the value of the portfolio at the beginning of the year minus the value of the portfolio at the end of the year. The results of many simulation trials are used to produce a complete credit loss distribution.

[16] The LGD is one minus the recovery rate.

[17] In this case, we assume that the recovery rate is the same at an extreme percentile of the default rate distribution and at the mean.

[18] See J. Hull, "Risk Management and Financial Institutions," Wiley, Fifth edition, 2018, Chapter 15 for more information on Basel II.

Table 6.4 Rating Transition Probabilities for a Borrower Rated B

Transition	Probability
"B" to "A"	0.05
Stays "B"	0.80
"B" to "C"	0.13
Defaults	0.02

Table 6.5 How the Sample from a Standard Normal Determines the Transitions in Table 6.4

Sample Range	Rating Transition
Less than −1.645	"B" to "A"
Between −1.645 and 1.036	Stays "B"
Between 1.036 and 2.054	"B" to "C"
Greater than 2.054	Defaults

We can illustrate the CreditMetrics model with a simple example. Suppose a bank uses four ratings: A, B, C, and default. As indicated in Table 6.4, a B-rated borrower has a 5% chance of being upgraded to A during the year, an 80% chance of staying at a B, a 13% chance of being downgraded to C, and a 2% chance of defaulting.

On each simulation trial, a number is sampled from a standard normal distribution to determine what happens to a given B-rated borrower.[19] Since $N^{-1}(0.05) = -1.645$, $N^{-1}(0.85) = 1.036$, and $N^{-1}(0.98) = 2.054$, the rules that are implemented are shown in Table 6.5. These rules ensure the transition probabilities in Table 6.4 apply.

As we have stressed throughout this chapter, the bank's borrowers generally do not default independently of each other. To reflect this, we need to sample from the distributions in a way that builds in a correlation between the samples. A factor model (like that described in Section 6.3) is usually used to define correlations between the normal distributions. The Monte Carlo simulation is therefore an implementation of the Gaussian copula model. The probability distribution of rating transitions for each borrower is transformed into a normal distribution, and the correlations are those between the normal distributions (not those between the rating transitions themselves).

The correlations between the returns on traded equities are often used to define the correlations used in CreditMetrics. This is an approach that can be justified using Merton's model.[20] Recall from Chapter 4 that this is a model where a company defaults when the market value of its assets falls below the book value of its debt.

One of the things distinguishing CreditMetrics from the Vasicek model and the model in Section 6.2 is that it considers the impact of rating changes as well as defaults. If the B-rated company in our example were downgraded to C in a simulation trial, the discount rate used at the end of the year to value the cash flows would be increased, and therefore the value of the loan would decline. The bank would incur a credit loss even though there was no default.[21]

6.6 RISK ALLOCATION

A result developed by a famous mathematician, Leonhard Euler, many years ago can be used to divide many of the risk measures used by risk managers into their component parts. Euler's result is concerned with what are termed homogeneous functions. These are functions, F, of a set of variables $x_1, x_2, \ldots, x_n$ having the property:

$$F(\lambda x_1, \lambda x_2, \ldots, \lambda x_n) = \lambda F(x_1, x_2, \ldots, x_n)$$

for a constant λ. Define

$$Q_i = x_i \frac{\Delta F_i}{\Delta x_i} \tag{6.8}$$

where Δx_i is a small change in variable i, and ΔF_i is the resultant small change in F. Here, Q_i is the ratio of ΔF_i to a proportional change, $\Delta x_i / x_i$, in x_i. Euler showed that in the limit, as the size of the Δx_i tends to zero:[22]

$$F = \sum_{i=1}^{n} Q_i$$

Many risk measures are homogeneous functions. Indeed, we learned in Chapter 1 that homogeneity is one of the properties of a coherent risk measure. If a portfolio is changed so that each

[19] The reason why we sample from a standard normal distribution rather than a simple uniform distribution between 0 and 1 will become apparent when we consider how default correlations are taken into account.

[20] See R. Merton, "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," *Journal of Finance*, 29 (1974): 449–470.

[21] As mentioned earlier, Basel II tries to take account of credit losses caused by a decline in creditworthiness during the year (but no default) with a "maturity adjustment" to the Vasicek model.

[22] To be precise, we should use calculus to define Q_i as

$$Q_i = x_i \frac{\partial F}{\partial x_i}$$

Table 6.6 Correlations Between Loan Losses

	Loan 1	Loan 2	Loan 3
Loan 1	1	0	0
Loan 2	0	1	0.7
Loan 3	0	0.7	1

position is multiplied by some constant λ, a risk measure is usually multiplied by λ. Euler's theorem therefore gives us a way of allocating a risk measure F that is a function of many different trades into its component parts.

In the context of credit risk, we can use Euler's theorem to determine the contribution of each loan in a portfolio to the overall risk measure. To illustrate this, let us take a very simple risk measure, standard deviation, and a portfolio consisting of three loans. We suppose that the losses from loans 1, 2, and 3 have standard deviations of 1.1, 0.9, and 0.9, respectively. The correlations between the losses are as shown in Table 6.6.

From Equation (6.1), the standard deviation of the total loss is

$$\sqrt{1.1^2 + 0.9^2 + 0.9^2 + 2 \times 0.7 \times 0.9 \times 0.9} = 1.99$$

Now suppose that the size of Loan 1 is increased by 1%. The standard deviation of the loss from Loan 1 increases from 1.1 to $1.1 \times 1.01 = 1.111$. The increase in the standard deviation of the loan portfolio is

$$\sqrt{1.111^2 + 0.9^2 + 0.9^2 + 2 \times 0.7 \times 0.9 \times 0.9}$$
$$- \sqrt{1.1^2 + 0.9^2 + 0.9^2 + 2 \times 0.7 \times 0.9 \times 0.9} = 0.006098$$

From Equation (6.8), $Q_1 = 0.006098/0.01 = 0.6098$.

Consider next the effect of increasing the size of Loan 2 by 1%. The standard deviation of the loss from Loan 2 increases from 0.9 to $0.9 \times 1.01 = 0.909$. The increase in the standard deviation of the loan portfolio is

$$\sqrt{1.1^2 + 0.909^2 + 0.9^2 + 2 \times 0.7 \times 0.909 \times 0.9}$$
$$- \sqrt{1.1^2 + 0.9^2 + 0.9^2 + 2 \times 0.7 \times 0.9 \times 0.9} = 0.006924$$

From Equation (6.8), $Q_2 = 0.006924/0.01 = 0.6924$.

Similarly, Q_3 also equals 0.6924 so that:

$$Q_1 + Q_2 + Q_3 = 0.6098 + 0.6924 + 0.6924 = 1.99$$

As predicted by Euler's theorem, we have divided the total loss of 1.99 into a contribution from the first loan, a contribution from the second loan, and a contribution from the third loan.[23]

[23] Note that the total of the three contributions (as we have calculated them) is not quite the same as the total portfolio standard deviation. The former is 1.9947, while the latter is 1.9910. In the limits, as the percentage changes that are considered are made smaller, the result becomes more exact.

Note that although Loan 1 has the greatest standard deviation, it has the lowest contribution to the total standard deviation. This is because it is uncorrelated with the other two loans and therefore adds less to the risk of the portfolio than they do.

6.7 DERIVATIVES

In this chapter, we have considered losses from loan portfolios. However, derivatives also give rise to credit risk. For example, if Company A buys an option from Company B, it is subject to the risk that B might default and therefore not provide a payout on the option when it is due. Similarly, if Company A enters an interest rate swap with Company B, it will potentially incur a credit loss if Company B defaults when the value of the swap to Company A is positive.

Equation (6.7) is used to calculate credit risk capital for derivatives as well as loans. But a problem in using Equation (6.7) for derivatives is that it is difficult to calculate EAD for a derivatives transaction. In the case of a loan, EAD is usually the amount that has been advanced (or is expected to be advanced) to the borrower. In the case of a derivative, the exposure varies from day to day as the value of the derivative changes. The Basel Committee's standard rules for determining EAD have solved this problem by setting the exposure at default for derivatives equal to the current exposure plus an add-on amount. The current exposure is the maximum amount that could be lost if the counterparty defaulted today. The add-on amount is an allowance for the possibility of the exposure getting worse by the time a default occurs.

Another consideration is that derivatives are subject to netting agreements. This means all outstanding derivatives with a counterparty may be considered a single derivative in the event that the counterparty defaults. This means that Equation (6.7) cannot be used on a transaction-by-transaction basis. Instead, it must be implemented on a counterparty-by-counterparty basis.

6.8 CHALLENGES

The assessment of credit risk requires many estimates (e.g., PD). Just as we distinguished between through-the-cycle and point-in-time for credit ratings in Chapter 4, we can distinguish between a through-the-cycle PD and a point-in-time PD. A through-the-cycle PD is an average PD over an economic cycle, whereas a point-in-time PD reflects current economic conditions.

Regulators require banks to estimate a through-the-cycle PD for regulatory capital purposes. However, point-in-time estimates might be more appropriate for internal purposes.[24] A further complication is that accounting standards (such as IFRS 9) require loans to be valued for accounting purposes. Expected losses (either over one year or during the loan's life, depending on the situation) must be calculated and subtracted from the principal amount of the loan. For this purpose, a point-in-time estimate is required. Banks are therefore faced with the problem of making both through-the-cycle estimates (to satisfy regulators) and point-in-time estimates (to satisfy their auditors).

Another required estimate is the recovery rate (or loss given default). As noted, the recovery rate is negatively correlated with the default rate. Thus, an economic downturn is doubly bad for credit risk because the default rate increases, and the recovery rate decreases.

A further estimate is exposure at default. This is the amount the borrower owes at the time of default. In the case of an overdraft facility or line of credit, this can be conservatively estimated as the borrowing limit assigned to the customer. (If the customer gets into financial difficulties, it is likely that the maximum amount will have been borrowed.) In a term loan, it is the expected principal amount during the year. For a portfolio of derivative transactions, a relatively complex calculation is required to determine the expected exposure during a year. A consideration might be what is termed *wrong-way risk*. This is the risk associated with the fact that a counterparty to a company may be more likely to default when the value of outstanding derivatives is negative to the counterparty (and therefore positive to the company).

Correlations are difficult to estimate. The Gaussian copula model is easy to use, but there is no guarantee it reflects just how bad loan losses would be in a one-in-a-thousand years scenario (or in the case of economic capital, an even more extreme scenario).

A final point to keep in mind is that credit risk is only one of many risks facing a bank. It also must worry about market risk, operational risk, liquidity risk, strategic risk, and so on. These risks tend to be handled by different groups within a bank, but they are not independent of each other. They interact and influence the bank's total required capital (both regulatory and economic).[25]

[24] Regulators feel that if they used point-in-time estimates for regulatory capital it would accentuate cycles. Banks would be less likely to lend during a low point in the economic cycle because capital requirements would be relatively high, and more inclined to do so during a high point in the cycle because regulatory requirements would be relatively low.

SUMMARY

If loans were independent of each other, there would be very little variation in the default rate experienced on a loan portfolio from one year to the next. In practice, as Table 6.1 shows, default rates do vary from year to year. Sometimes there is credit contagion, where a default by one company leads to a default by other companies that are linked to the first company in some way. As an approximation, it can be argued that, conditional on the performance of the economy, companies default independently of each other. But ups and downs in the economy create default correlation.

We have presented three models of default risk. The first model allows the standard deviation of the loss on a loan portfolio to be calculated from the characteristics of the loans themselves and the correlations between them. However, risk managers are generally interested in more than just a standard deviation of a loan portfolio. They want to know high percentiles of the credit loss distribution. Unfortunately, there is no easy way to calculate these from the standard deviation.

The second model, the Vasicek model, is the model used by regulators to determine the capital banks are required to hold for credit risk. Its advantage is that for a large portfolio, high percentiles of the loss distribution can be calculated analytically. Given an estimate of the default probability for loans in a portfolio, it is possible to immediately estimate the 99.9 percentile (or an even higher percentile) of the loss distribution.

The third model, CreditMetrics, is often used by financial institutions to determine economic capital. Economic capital is a bank's own estimate of the capital it requires and may be based on a higher confidence level than regulatory capital. Typically, CreditMetrics involves time-consuming Monte Carlo simulations.

The famous Swiss mathematician Leonhardt Euler died in 1783. If he were still alive, he might be pleased to see how some of his theoretical results are used in risk management. He provided a way in which the total risk from a number of positions can be divided into a risk associated with each position so that the sum of the position risks equals the total risk.

[25] Enterprise risk management is the term used to describe a holistic approach to risk management, where all risks are considered in conjunction with a company's risk appetite and its risk culture.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

QUESTIONS

Short Concept Questions

6.1 What is the difference between economic capital and regulatory capital?

6.2 How does a bank decide how much equity capital it needs?

6.3 Which model is used by the Basel Committee in its internal ratings-based approach for calculating credit risk?

6.4 What is meant by (a) expected loss and (b) unexpected loss?

6.5 Is the correlation between recovery rates and default rates positive or negative?

6.6 As the default correlation in a loan portfolio increases, does a bank need more or less capital?

6.7 How can the Gaussian copula model be used to define the joint probability distribution of two random variables?

6.8 Explain what PD, LGD, and EAD mean in the determination of regulatory capital.

6.9 What does it mean to assert that a risk measure is homogeneous?

6.10 Give two reasons why it is more difficult to quantify credit risk for derivatives than for loans.

Practice Questions

6.11 A USD 1 million loan has a probability of 0.5% of defaulting in a year. The recovery rate is estimated to be 40%. What is the expected credit loss and the standard deviation of the credit loss?

6.12 Suppose that there are three USD 1 million loans like the one in the previous question. The correlation between any pair of the loans is 0.2. What is the mean and standard deviation of the portfolio credit loss?

6.13 In Chapter 4, we saw that statistics produced by rating agencies show that a BBB-rated company has a probability of 0.18% of defaulting in one year, and an A-rated company has a probability of 0.06% of defaulting in one year. A bank currently has a BBB credit rating. How should it determine its economic capital if (a) its objective is to avoid being downgraded and (b) its objective is to be upgraded to a rating of A?

6.14 A bank has a USD 100 million portfolio of loans with a PD of 0.75%. What is the 99.9 percentile of the default rate given by the Vasicek model? Assume a correlation parameter of 0.2.

6.15 In the situation considered in the previous question, the recovery rate in the event of a default is 30%. What is the required regulatory capital?

6.16 For economic capital purposes, a bank uses a 99.97% confidence level. What is the economic capital in the situation considered in the previous two questions? Continue to base your answers on the Vasicek model.

6.17 From Table 6.1, it can be seen that defaults are greatest during and after recessions. How does the Vasicek model account for economic conditions?

6.18 X and Y are variables that have uniform distributions between 0 and 1 (a uniform distribution is a distribution where all values in a certain range are equally likely). A Gaussian copula model is used to define a correlation between them. The correlation parameter is 0.25. How would you determine the probability that both are less than 0.5?

6.19 The standard deviations of the losses on three loans are all 1.5. Loan 1 and Loan 2 are uncorrelated. There is a correlation of 0.2 between Loan 1 and Loan 3, and a correlation of 0.5 between Loan 2 and Loan 3. Calculate the standard deviation of the loss on the portfolio.

6.20 In the previous question, use Euler's result to determine the contribution of each loan to the total standard deviation.

ANSWERS

Short Concept Questions

6.1 Economic capital is a bank's own estimate of the capital it requires. Regulatory capital is the capital that banks are required to keep by regulators.

6.2 It tries to ensure that it will not run out of capital in an extremely bad year.

6.3 The Vasicek model

6.4 The expected loss is the loss in an average year. The unexpected loss is a high percentile of the loss distribution minus the expected loss.

6.5 It is negative. When the default rate is high (or low) the recovery rate tends to be low (or high).

6.6 As default correlation increases, more capital is required.

6.7 Each variable is mapped to a standard normal distribution on a percentile-to-percentile basis. The two normal distributions are then assumed to have a bivariate normal distribution with a particular correlation.

6.8 PD is probability of default, LGD is loss given default (= one minus the recovery rate), and EAD is exposure at default.

6.9 It means that when the sizes of all the positions are multiplied by a constant amount, the risk measure itself is multiplied by that amount.

6.10 The exposure at default for a derivative is much less certain than it is for a loan. Also, derivatives are subject to netting agreements, which mean that all outstanding derivatives are treated as a single derivative in the event of a default.

Solved Problems

6.11 The expected loss in USD is $0.005 \times 1 \times (1 - 0.4) = 0.003$. This is USD 3,000. The variance of the loss is $0.005 \times 0.6^2 - (0.005 \times 0.6)^2 = 0.001791$. The standard deviation is the square root of this, or USD 0.04232 million. This is USD 42,320.

6.12 The expected loss on the three loans is three times the expected loss on one loan or USD 9,000. The variance of the portfolio loss is

$$3 \times 0.001791 + 6 \times 0.001791 \times 0.2 = 0.007522$$

The standard deviation of the portfolio loss is the square root of this or USD 0.086731. This is USD 86,731. (Note that the standard deviation is less than three times the standard deviation calculated for one loan because there are some diversification benefits.)

6.13 In (a), the economic capital should be the unexpected loss calculated using a confidence level of 99.82%. In (b), the economic capital should be the unexpected loss calculated using a confidence level of 99.94%.

6.14 From Equation (6.5), the default rate is

$$N\left(\frac{N^{-1}(0.0075) + \sqrt{0.2}N^{-1}(0.999)}{\sqrt{1 - 0.2}}\right) = 0.1201$$

or 12.01%.

6.15 From Equation (6.6), this is in USD million:

$$(0.1201 - 0.0075) \times 100 \times 0.7 = 7.88$$

6.16 The 99.97 percentile of the default rate is

$$N\left(\frac{N^{-1}(0.0075) + \sqrt{0.2}N^{-1}(0.9997)}{\sqrt{1 - 0.2}}\right) = 0.1578$$

The capital in USD million is

$$(0.1578 - 0.0075) \times 100 \times 0.7 = 10.52$$

6.17 In the Vasicek model, there is a normally distributed factor F describing the health of the economy. Conditional on F, obligors default independently of each other. The value of F determines the default rate on a large portfolio.

6.18 In the Gaussian copula model, the 0.5 values for X is transformed to the zero value of a standard normal distribution. The same is true for Y. The required probability is the probability that the first variable is less than zero, and the second variable is less than zero in a bivariate normal probability distribution where the coefficient of correlation is 0.25 (it can be shown that this is about 0.29).

6.19 The variance is

$$1.5^2 + 1.5^2 + 1.5^2 + 2 \times 0.2 \times 1.5 \times 1.5 + 2 \times 0.5 \times 1.5 \times 1.5 = 9.9$$

The standard deviation is the square root of this or 3.146.

6.20 Increasing Loan 1 by 1% increases the standard deviation by:

$$\sqrt{1.515^2 + 1.5^2 + 1.5^2 + 2 \times 0.2 \times 1.515 \times 1.5 + 2 \times 0.5 \times 1.5 \times 1.5}$$
$$-\sqrt{1.5^2 + 1.5^2 + 1.5^2 + 2 \times 0.2 \times 1.5 \times 1.5 + 2 \times 0.5 \times 1.5 \times 1.5}$$
$$= 0.008605$$

0.008605/0.01 or 0.86 of the standard deviation is therefore allocated to Loan 1.

Increasing Loan 2 by 1% increases the standard deviation by:

$$\sqrt{1.5^2 + 1.515^2 + 1.5^2 + 2 \times 0.2 \times 1.5 \times 1.5 + 2 \times 0.5 \times 1.515 \times 1.5}$$
$$-\sqrt{1.5^2 + 1.5^2 + 1.5^2 + 2 \times 0.2 \times 1.5 \times 1.5 + 2 \times 0.5 \times 1.5 \times 1.5}$$
$$= 0.01074$$

0.01074/0.01 or 1.07 is therefore allocated to Loan 2.

Increasing Loan 3 by 1% increases the standard deviation by:

$$\sqrt{1.5^2 + 1.5^2 + 1.515^2 + 2 \times 0.2 \times 1.5 \times 1.515 + 2 \times 0.5 \times 1.5 \times 1.515}$$
$$-\sqrt{1.5^2 + 1.5^2 + 1.5^2 + 2 \times 0.2 \times 1.5 \times 1.5 + 2 \times 0.5 \times 1.5 \times 1.5}$$
$$= 0.01217$$

0.01217/0.01 or 1.22 is therefore allocated to Loan 3.

7

Operational Risk

Learning Objectives

After completing this reading, you should be able to:

- Describe the different categories of operational risk and explain how each type of risk can arise.
- Compare the basic indicator approach, the standardized approach, and the advanced measurement approach for calculating operational risk regulatory capital.
- Describe the standardized measurement approach and explain the reasons for its introduction by the Basel Committee.
- Explain how a loss distribution is derived from an appropriate loss frequency distribution and loss severity distribution using Monte Carlo simulation.
- Describe the common data issues that can introduce inaccuracies and biases in the estimation of loss frequency and severity distributions.
- Describe how to use scenario analysis in instances when data are scarce.
- Describe how to identify causal relationships and how to use Risk and Control Self-Assessment (RCSA), Key Risk Indicators (KRIs), and education to understand and manage operational risks.
- Describe the allocation of operational risk capital to business units.
- Explain how to use the power law to measure operational risk.
- Explain how the moral hazard and adverse selection problems faced by insurance companies relate to insurance against operational risk.

Understanding operational risk has become increasingly important for banks, insurance companies, and other financial institutions. There are many ways operational risk can be defined. It is sometimes defined very broadly as any risk that is not a market risk or a credit risk. A much narrower definition would be that it consists of risks arising from operational mistakes; this would include the risk that a bank transaction is processed incorrectly, but it would not include the risk of fraud, cyberattacks, or damage to physical assets.

Operational risk has been defined by the Basel Committee as:

> *The risk of loss resulting from inadequate or failed internal processes, people, and systems or from external events.*[1]

The International Association of Insurance Supervisors defines operational risk similarly as:

> *The risk of adverse change in the value of capital resources resulting from operational events such as inadequacy or failure of internal systems, personnel, procedures or controls, as well as external events.*

These definitions include the risks arising from computer hacking, fines from regulatory agencies, litigation, rogue traders, terrorism, systems failures, and so on. However, they do not include strategic risks or reputational risks.

Seven categories of operational risk have been identified by the Basel Committee.[2]

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, and insider trading on an employee's own account.
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 work place safety*: Acts inconsistent with employment, health or safety laws or agreements, or which result in payment of personal injury claims, or claims relating to diversity or discrimination issues. Examples include workers 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 client's accounts, non-client counterparty misperformance, and vendor disputes.

Operational risk is much more difficult to quantify than market or credit risk. In the case of market risk, risk factor volatilities can be estimated from historical data so that risk measures (such as VaR) can be used to quantify potential losses. In the case of credit risk, rating agencies and internal bank records can provide data on expected and unexpected losses. For operational risk losses, however, there is relatively little data. What is the probability of a cyberattack destroying bank records? What is the probability of a rogue trader loss? How much would the bank lose if these events occurred? Many serious operational risks facing financial institutions involve rare and novel events.

This chapter looks at the way regulatory and economic capital is calculated for operational risk. It also looks at how financial institutions can take proactive steps to both reduce the chance of adverse events happening as well as minimize the severity of the outcomes when they do happen.

7.1 LARGE RISKS

We start this chapter by reviewing three large operational risks faced by financial institutions.

Cyber Risks

Cyber risk is one of the largest risks faced by financial institutions. They and their clients have benefited from the development of credit and debit cards, online banking, mobile wallets,

[1] https://www.bis.org/publ/bcbs195.pdf

[2] See Basel Committee on Bank Supervision, "Sound Practices for the Management and Supervision of Operational Risk," February 2003.

electronic funds transfer, and so on. However, these innovations have also created opportunities for hackers. Cyberattacks are proving to be very expensive for businesses.[3] These attacks are increasing in sophistication and severity with each passing year. Threats can come in the form of individual hackers, nation states, organized crime, and even insiders. Defenses developed against cyber threats include user account controls, cryptography, intruder detection software, and firewalls.

Cyber crime includes the destruction of data, theft of money, theft of intellectual property, theft of personal and financial data, embezzlement, fraud, and so on. For example, a 2013 cyberattack at Yahoo saw a data breach that included the names, email addresses, telephone numbers, security questions and answers, dates of birth, and passwords for three billion user accounts. Equifax, a large consumer credit reporting agency, reported a cyberattack in 2017 affecting 143 million people in the United States. Sensitive information such as social security numbers and driver license numbers were obtained by the hackers.

Large corporations are under continuous attack by cyber criminals. Most attacks are unsuccessful, but the number of reported successful attacks appears to be rising each year. In addition, there may be many successful cyberattacks that go unreported.

Financial institutions are targeted in many ways. Most readers will be familiar with a common hacking practice called phishing. While phishing can come in many forms, a common situation involves a hacker targeting a financial institution's customers with an email asking them to confirm account information. If a customer complies, the criminal gains access to the customer's account. Sometimes the customer is tricked into installing malicious software that allows the hacker to capture sensitive information as it is typed.

A more serious threat is to the financial institution itself. If a hacker can gain access to a financial institution's systems, he or she can obtain client information, delete records, enter false transactions, embezzle funds, and so on. The dangers here were illustrated in the March 2016 hacking of the Central Bank of Bangladesh. The hackers found multiple entry points into the bank's network and planned to embezzle over USD 1 billion through a series of international transactions. However, it is reported that because of a data entry mistake, they obtained only USD 80 million (which is still an embarrassingly large sum).

All companies should accept that, however good their defenses are, they are liable to be hacked in the next few years. They should have plans that can be implemented at short notice to deal with attacks of different severities. In some instances, an extreme response to an attack, such as delaying the acceptance of new transactions for a few days, might be necessary.

Compliance Risks

Compliance risk is another operational risk facing financial institutions. This is the risk that an organization will incur fines or other penalties because it knowingly or unknowingly fails to act in accordance with industry laws and regulations, internal policies, or prescribed best practices. Activities such as money laundering, terrorism financing, and assisting clients with tax evasion can all lead to big penalties.

A well-known example of compliance risk is Volkswagen's failure to comply with U.S. emissions standards by cheating during emissions testing. This led to a fine of about USD 2.8 billion.

One example of compliance risk in the financial sector is the USD 1.9 billion penalty paid by HSBC in 2012. In the years preceding the fine, the bank did not implement anti-money laundering programs for its Mexican branches. As a result, Mexican drug traffickers were able to illegally deposit large sums of money in cash. HSBC eventually reached a deferred prosecution agreement with the United States Department of Justice, which required it to pay a large fine and retain an independent compliance monitor.

Another example of financial sector compliance risk comes from French bank BNP Paribas. In 2014, it was announced the bank would pay USD 8.9 billion (roughly one-year's profit) to the United States government for moving dollar-denominated transactions through the U.S. banking system on behalf of Sudanese, Iranian, and Cuban parties. These transactions occurred despite the fact all three countries were subject to economic sanctions by the U.S. government that banned the transactions. In addition to paying the fine, BNP Paribas was also banned from conducting certain U.S. transactions for a year.

Regulatory infractions can result from a small part of a large company`s global activities. However, they can be very expensive (both in terms of fines and loss of reputation). It is important for financial institutions to have systems in place to ensure that they are following all applicable laws and regulations. In this regard, technology can be of help. For example, some banks have developed systems designed to detect suspicious requests to open accounts or transfer funds in real time.[4]

[3] Forbes estimates the direct and indirect cost to all businesses to be as high as USD 6 trillion per year. See "The True Cost of Cybercrime for Businesses," Forbes, July 2017.

[4] Many readers will have received requests from a bank to verify a transaction that it has flagged as suspicious. This is a simple example of the types of applications being talked about here.

Rogue Trader Risk

Rogue trader risk is the risk that an employee will take unauthorized actions resulting in large losses. One of the most notorious incidents involved Barings Bank trader Nick Leeson. His job was to do relatively low-risk trades from the firm's Singapore office. Due to flaws in the Barings systems, however, he found a way of taking large risks and hiding losses in a secret account. His attempts to recoup losses led to even more losses (which exceeded USD 1 billion) and he was forced to flee Singapore to avoid prosecution (leaving a note saying he was sorry). Leeson was eventually returned to Singapore, prosecuted, and received a prison sentence. Barings Bank, which had been in existence for 200 years, was forced into bankruptcy.

Another large loss occurred at the Sociéte Générale (SocGen). Ostensibly, trader Jérôme Kerviel was tasked with finding arbitrage opportunities in equity indices, such as the German DAX, the French CAC 40, and the Euro Stoxx 50. These might arise if a futures contract on an index was trading at different prices on two exchanges, or at a price inconsistent with the prices of the underlying shares. However, Kerviel found a way of speculating while appearing to arbitrage. He took big positions and created fictitious trades to make it appear as if he was hedged. In January 2008, Kerviel's unauthorized trading was uncovered and SocGen lost EUR 4.9 billion when his positions were closed out.

Other rogue trading incidents include a USD 2.3 billion loss at UBS in 2011 and a USD 700 million loss at Allied Irish Bank in 2002. The common theme among these losses is that a single trader working for a bank was able to take huge risks without the firm's knowledge or authorization. These actions were hidden via the creation of fictitious offsetting trades, or in some other way. In every case, the trader hoped that by continuing to speculate, losses would be offset, and the unauthorized trading would have then been forgiven. (Indeed, there are almost certainly cases of unauthorized trading that we are unaware of because a trader's doubling-down strategy successfully reversed losses.)

One thing a bank can do to protect itself is to ensure that the front office (which is responsible for trading) is totally independent of the back office (which is responsible for record keeping and verifying transactions). A more difficult issue is the way in which unauthorized trading is treated when it is uncovered. If a trader conducts unauthorized trading and takes a loss, there are likely to be unpleasant consequences for the trader. But what if the trader makes a profit? It is then tempting to ignore the violations. This is short-sighted, however, because it leads to a culture where risk limits are not taken seriously. This in turn paves the way for disaster.

7.2 BASEL II REGULATIONS

As explained in earlier chapters, the Basel Committee on Banking Supervision develops global regulations, which are then implemented by bank supervisors in each member country. In 1999, it issued an early draft of a regulation that became known as Basel II.[5] In part, this was a revision of the methods for calculating credit risk capital (which were discussed in Chapter 6). A surprise inclusion, however, was a clear indication that regulators were planning to require banks to hold capital for operational risk in addition to the capital already required for market risk and credit risk.

Many risk managers deemed capital requirements for operational risk to be unworkable because of the difficulty in quantifying operational risk. However, the Basel Committee recognized that many of the large losses experienced by banks were operational risk losses, not market risk or credit risk losses. Even if operational risks could not be quantified precisely, they considered it important for banks to devote more resources toward managing them.

There has been a parallel development in the regulation of insurance companies. Solvency II, the European Union's regulatory framework for insurance companies issued in 2016, requires capital to be held for operational risk. As mentioned earlier, insurance regulators use a definition of operational risk similar to that used by Basel II.

The result of these regulations has been that operational risk management is given much more emphasis within banks and insurance companies. Operational risk managers are important members of the risk management team at financial institutions. They must understand where losses might occur, which losses should be insured against, and how losses can be mitigated. In addition to requiring capital for operational risk, regulators have also produced guidelines on how operational risk should be managed and have indicated a desire to see evidence that the guidelines are being followed.[6]

It has not been easy for regulators to come up with rules for determining operational risk capital. The final Basel II rules for banks had three approaches[7]:

1. The basic indicator approach,
2. The standardized approach, and
3. The advanced measurement approach (AMA).

[5] See Basel Committee on Bank Supervision, "A New Capital Adequacy Framework," June 1999.

[6] See Basel Committee on Bank Supervision, "Sound Practices for the Management and Supervision of Operational Risk," February 2013.

[7] See Basel Committee on Bank Supervision, "International Convergence of Capital Measurement and Capital Standards," June 2006.

Table 7.1 Capital as a Percentage of Gross Income for Different Business Lines in the Standardized Approach

Business Line	Capital (% of Gross Income)
Corporate finance	18%
Trading and sales	18%
Retail banking	12%
Commercial banking	15%
Payment and settlement	18%
Agency services	15%
Asset management	12%
Retail brokerage	12%

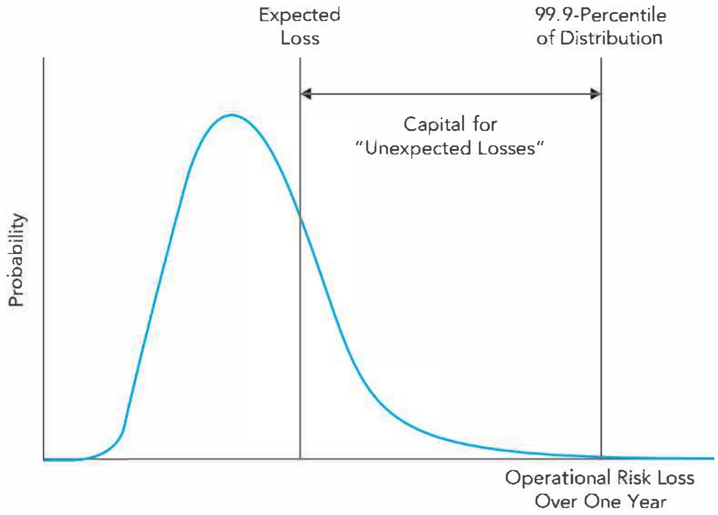


Figure 7.1 The AMA model.

Banks typically started by using the basic indicator approach. They then needed to satisfy several criteria to be permitted to use the standardized approach. After that, they could use AMA by satisfying further criteria. The first two approaches were quite simple. The third approach was quite complicated.

In the basic indicator approach, operational risk capital was set equal to 15% of the three-year average annual gross income. Gross income is defined as:

$$\text{Interest earned} - \text{interest paid} + \text{non-interest income}$$

The standardized approach was similar, except that separate calculations are carried out by each business line and the percentage applied to gross income varies across business lines. The percentages are in Table 7.1.

The AMA in Basel II was much more complicated than the other two approaches. Banks were required to treat operational risk like credit risk and set capital equal to the 99.9 percentile of the loss distribution minus the expected operational loss. The model is illustrated in Figure 7.1 (which is similar to Figure 6.2 for credit risk).

Under the AMA approach, banks were required to consider every combination of the eight business lines in Table 7.1 and the seven risk types mentioned in the introduction. For each of the 56 ($= 7 \times 8$) combinations, they had to estimate the 99.9 percentile of the one-year loss. These estimates were then aggregated to determine the total capital requirement.[8]

The Basel Committee has now abandoned AMA and is replacing all three of the approaches in Basel II with a new standardized approach (which will be discussed in the next section). However, we will later examine some key aspects of the AMA approach because many banks still use it as part of their economic capital determinations. Specifically, the calculation of economic capital requires a probability distribution for the one-year loss and uses the model in Figure 7.1 (but usually with a higher percentile than 99.9%).

7.3 REVISION TO BASEL II

The AMA operational risk methodology we have described has proved useful in prompting financial institutions to think more about the operational risks they face. However, bank regulators have found the approach unsatisfactory due to the high degree of variation in the calculations carried out by different banks. Two banks presented with the same data were liable to come up with quite different capital requirements under AMA.

The Basel Committee therefore announced in March 2016 its intention to replace all previous approaches for determining operational risk capital with a new approach: the standardized measurement approach (SMA).[9] The SMA first defines a quantity known as the Business Indicator (BI). BI is like gross income, but it is designed to be a more relevant measure of bank size. For example, items such as trading losses and operating expenses, which reduce gross income, are treated differently so that they increase BI.

The BI Component for a bank is calculated from the BI using a piecewise linear relationship[10]. A loss component is then calculated as:

$$7X + 7Y + 5Z$$

Here, X, Y, and Z are estimates of the average annual loss from operational risk over the previous ten years. The quantity X

[8] See Basel Committee on Bank Supervision, "Supervisory Guidelines and the Advanced Measurement Approach," June 2011

[9] See Basel Committee on Bank Supervision, "Standardized Approach to Operational Risk: Consultative Document," March 2016.

[10] A piecewise function is composed of two or more sub-functions, each applicable across a certain interval of the function's domain.

includes all losses, Y includes only losses greater than EUR 10 million, and Z includes only losses greater than EUR 100 million. The calculations are designed so that the loss component and the BI Component are equal for an average bank. The Basel Committee provides a formula for calculating required capital from the loss component and the BI component.[11]

7.4 DETERMINING THE LOSS DISTRIBUTION

Economic capital calculations require a distribution (like that in Figure 7.1) for several categories of operational risk losses and the combined results. The key determinants of an operational risk loss distribution are

- *Average Loss frequency*: the average number of times in a year that large losses occur, and
- *Loss severity*: the probability distribution of the size of each loss.

Loss Frequency

A Poisson distribution is often assumed for loss frequency. This is a distribution of the number of events in a certain time if the events occur at a certain rate and are independent of each other. If the expected number of losses in a year is λ, the probability of n losses during the year given by the Poisson distribution is[12]

$$\frac{e^{-\lambda}\lambda^n}{n!}$$

Table 7.2 uses the Poisson distribution to give the probability for the number of losses in a year when the average number of losses in the year is 2, 4, and 6.

Table 7.2 Probability Distribution for Number of Losses in a Year for Three Different Values of Average Loss Frequency Parameter, λ

No. of Losses	Average Loss Frequency 2	4	6
0	0.135	0.018	0.002
1	0.271	0.073	0.015
2	0.271	0.147	0.045
3	0.180	0.195	0.089
4	0.090	0.195	0.134
5	0.036	0.156	0.161
6	0.012	0.104	0.161
7	0.003	0.060	0.138
8	0.001	0.030	0.103
9	0.000	0.013	0.069
10	0.000	0.005	0.041
11	0.000	0.002	0.023
12	0.000	0.001	0.011
13	0.000	0.000	0.005
14	0.000	0.000	0.002
15	0.000	0.000	0.001

Loss Severity

The mean and standard deviation of the loss severity is often fitted to a lognormal distribution. This is a distribution where the natural logarithm of the variable is normal. Suppose the mean and standard deviation of the loss size are estimated to be μ and σ, respectively. Under the lognormal assumption, the mean of the logarithm of the loss size is[13]

$$m = \ln\left(\frac{\mu}{\sqrt{1 + w}}\right)$$

and the variance of the logarithm of the loss size is

$$s^2 = \ln(1 + w)$$

where $w = (\sigma/\mu)^2$.

For example, if the mean and standard deviation of the loss size are estimated (in USD million) as 80 and 40, then $w = 0.5^2 = 0.25$. The logarithm of the loss size therefore has a mean of:

$$\ln\left(\frac{80}{\sqrt{1.25}}\right) = 4.27$$

and a variance of $\ln(1.25) = 0.223$.

Monte Carlo Simulation

Once λ, μ, and σ have been estimated, a Monte Carlo simulation can be used to determine the probability distribution of the loss. The general approach is illustrated in Figure 7.2.

The steps in the procedure are as follows.

Step 1: Sample from the Poisson distribution to determine the number of loss events ($= n$) in a year.

[11] See J. Hull, "Risk Management and Financial Institutions," Fifth edition, Wiley, 2018 for more information on the SMA.

[12] $n!$ is referred to as "n factorial" and defined as $n! = n \times (n-1) \times (n-2) \times \ldots \times 2 \times 1$.

[13] To prove these results, note that from the properties of the lognormal distribution we require $\mu = \exp\left(m + \frac{s^2}{2}\right)$ and $\sigma^2 = [\exp(s^2) - 1]\mu^2$

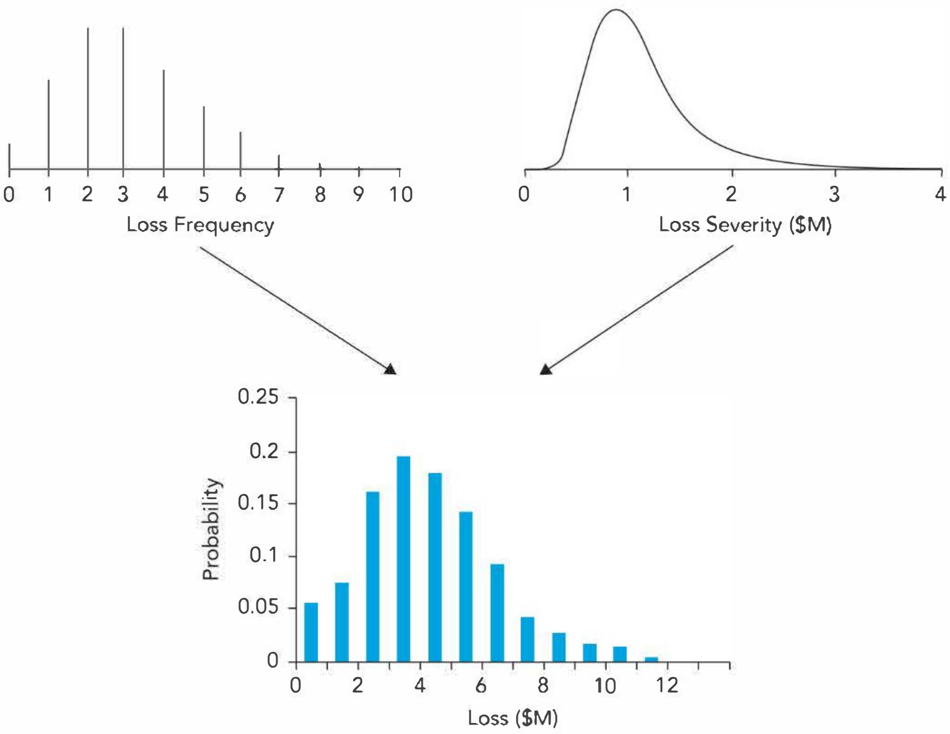


Figure 7.2 **Determination of loss distribution from loss frequency and loss severity.**

Step 2: Sample *n* times from the lognormal distribution of the loss size for each of the *n* loss events.

Step 3: Sum the *n* loss sizes to determine the total loss.

Step 4: Repeat steps 1 to 3 many times.

For step 1, we can sample the percentile of the Poisson distribution as a random number between zero and one. For example, suppose that the average loss frequency is 4 and the random number sampled is 0.31. From Table 7.2, this corresponds to three loss events. This is because the cumulative probability for two loss events or less is 0.238 (= 0.018 + 0.073 + 0.147) and for three loss events or less is 0.433 (= 0.018 + 0.073 + 0.147 + 0.195). The sampled number of 0.31 lies between these two cumulative probabilities.

Suppose the loss size has mean 80 and standard deviation 40 (as in the example above). For step 2, we sample three times from a normal distribution with mean 4.27 and variance 0.223. If the numbers sampled are 4.1, 5.1, and 4.4, the three losses are

$$e^{4.1} = 60.34$$
$$e^{5.1} = 164.02$$
$$e^{4.4} = 81.45$$

Step 3 gives the total loss on that simulation trial of 305.81 (= 60.34 + 164.02 + 81.45).

By carrying out many Monte Carlo simulation trials such as this, we obtain a probability distribution for the total loss from which the required percentile can be calculated.

Estimation Procedures

Estimating the loss frequency and loss severity for a category of losses involves a combination of data and subjective judgement. Loss frequency should be estimated either from a bank's own data or subjectively by operational risk professionals after careful consideration of the controls in place. We will examine subjective estimation further in the scenario analysis section later on.

When loss severity cannot be estimated from a financial institution's own data, the losses experienced by other financial institutions can sometimes be used as a guide. Mechanisms for sharing loss data between banks have been developed. Additionally, data vendor services (such as Factiva and Lexis-Nexis) can be useful in providing data on publicly reported losses experienced by other banks.

Potential Biases

Data from data vendors can potentially be biased because only large losses are usually reported. If the data from a vendor is used in a direct way to determine the loss severity distribution, the distribution is likely to be biased toward large losses. This bias can be avoided if the data is used to determine only relative loss severity. If data from a vendor indicates Loss Type A (on which a bank has no data) is on average twice as severe as Loss Type B (on which the bank does have data), the bank might assume that the mean loss for Loss Type A is twice that calculated using its own data for Loss Type B. Similarly, if vendor data indicates the standard deviation for Loss Type A is 50% greater than that for Loss Type B, the bank might assume the standard deviation of losses for Loss Type A is 50% greater than the standard deviation of losses for Loss Type B calculated using its own data.

Another potential bias concerns the size of a loss. Suppose Bank B has revenues of USD 20 billion and experiences a loss of USD 300 million. Bank A, with revenues of USD 10 billion, is using this loss event to estimate the severity of a similar loss it might incur.[14] Bank A's loss would most likely not be as large as USD 300 million because it is a smaller bank than Bank B. But it would be too optimistic to estimate its loss would be half of that of Bank B (USD 150 million). Shih et al. (2000) use vendor data to estimate a model of the form:[15]

$$\text{Estimated Loss for Bank A} = \text{Observed Loss for Bank B} \times \left(\frac{\text{Bank A Revenue}}{\text{Bank B Revenue}}\right)^{\beta}$$

They find that $\beta = 0.23$ gives a good fit. The loss for Bank A in our example would therefore be

$$300 \times \left(\frac{10}{20}\right)^{0.23} = 256$$

or USD 256 million. It is also important to adjust loss severity estimates for inflation. A loss of a certain size observed ten years ago can be expected to be larger if the same set of circumstances repeat themselves.

Scenario Analysis

Financial institutions also use scenario analysis to estimate loss frequencies and loss severities. It is particularly useful for loss events with a low frequency but high severity. These are the important loss events because they tend to determine the extreme tails of the loss distribution.

The objective for this approach is to list these events and generate a scenario for each one. Some of the scenarios might come from a financial institution's own experience, some might be based on the known experience of other banks, and some might be hypothetical scenarios generated by risk management professionals. Sometimes consultants are used to assist in generating scenarios.

For each scenario, loss frequency and loss severity estimates are made. Monte Carlo simulations (like the one illustrated in Figure 7.2) are used to determine a probability distribution for total loss across different categories of losses. The estimates are usually made by a committee of operational risk experts. The loss frequency estimate should reflect the controls in place at the financial institution and the type of business it is doing.

Estimating the probability of events that happen infrequently is difficult. One approach is to specify several categories and ask operational risk experts to assign each loss to a category. The categories could be

- Scenario happens once every 1,000 years on average ($\lambda = 0.001$),
- Scenario happens once every 100 years on average ($\lambda = 0.01$),
- Scenario happens once every 50 years on average ($\lambda = 0.02$),
- Scenario happens once every ten years on average ($\lambda = 0.1$), and
- Scenario happens once every five years on average ($\lambda = 0.2$).

Operational risk experts must also estimate loss severity. Rather than estimate the mean and standard deviation, it might be more appropriate to ask for estimates of the 1 percentile to 99 percentile range of the loss distribution. These estimates can be made to fit to a lognormal distribution. For example, suppose that 20 and 200 are the 1 percentile and 99 percentile of the loss (respectively). Then $\ln(20) = 2.996$ and $\ln(200) = 5.298$ are the 1 and 99 percentiles for the logarithm of the loss distribution (respectively). From this, it follows that the logarithm of the loss distribution has a mean of $(2.996 + 5.298)/2 = 4.147$ and a standard deviation of $(5.298 - 4.147)/2.326 = 0.49$.[16]

The key point here is that scenario analysis considers losses that have never been experienced by a financial institution yet could happen in the future. Managerial judgement is used to assess loss frequency and loss severity. Hopefully, this leads to an active discussion about how such loss events can occur. Scenario analysis can help firms form strategies for responding to a loss event and/or reducing the probability of it happening in the first place.

Allocation of Economic Capital

Economic capital is allocated to business units so that a return on capital can be calculated. The procedure for allocating credit risk capital was discussed in Section 6.6. A similar procedure can (in principle) be used for allocating operational risk capital.

The allocation of operational risk capital provides an incentive for a business unit manager to reduce operational risk. If the business unit manager can show that he or she has successfully reduced either loss frequency or loss severity, less capital will be allocated to the business unit. The unit's return on capital will then improve and the manager can hope for a bigger bonus. At the very least, the allocation process should sensitize the manager to the importance of operational risk.

[14] We define the revenue for a bank as its gross income (see Section 7.2).

[15] See J. Shih, A. Samad-Khan and P. Medapa, "Is Size of an Operational Loss Related to Firm Size," *Operational Risk Magazine*, 2, 1 (January 2000). The model is estimated by taking logarithms of both sides of the equation.

[16] The mean of a normal distribution is the average of the 1 and 99-percentile points. The 99-percentile of a normal distribution is 2.326 standard deviations above the mean.

Note that it is not always optimal to reduce operational risk. Some level of operational risk is inevitable in any business unit, and any decision to reduce operational risk by increasing operating costs should be justified with a cost-benefit analysis.

Power Law

As we saw in Chapter 6, economic capital is often calculated with very high confidence levels. For some probability distributions occurring in nature, it has been observed that a result known as the power law holds. If v is the value of a random variable and x is a high value of v, then the power law holds it is approximately true that:

$$\Pr(v > x) \approx Kx^{-\alpha} \tag{7.1}$$

where Pr denotes probability, and K and α are parameters.

The power law describes how fat the right tail of the probability distribution of v is. The parameters K and α depend on the variable being considered. K is a scale parameter, while α reflects the fatness of the distribution's right tail. As the parameter α decreases, this tail becomes fatter.

The power law only describes the right tail of the distribution (not the whole distribution). That is why Equation (7.1) is approximately true only for high values of x (values of x that are well into the right tail of the distribution of v). The power law is based on the work of Polish mathematician B.V. Gnedenko.[17] He showed that the tails of a wide range of distributions share the common properties indicated in Equation (7.1). To be mathematically correct, we should say Equation (7.1) is true for a wide range of distributions in the limit as x tends to infinity. In practice, it is usually assumed to be true for values of x that are in the top 5% of the distribution.

Gnedenko's result has been shown to be true for a wide range of distributions, such as:

- The incomes of individuals,
- The magnitude of earthquakes,
- The sizes of cities as measured by population,
- The sizes of corporations,
- The trading volume of a stock,
- The occurrence of a word in text, and
- The number of hits to a website.

As a rough statement, we can say that the power law holds for probability distributions of random variables which are the result of aggregating many independent random effects in some manner.[18]

Work by de Fontnouvelle et al. (2003) suggests the power law holds for operational risk losses.[19] This can be useful in some circumstances. For example, supposed K and α are estimated as 10,000 and 3 (respectively).[20] Suppose further that we are interested in estimating the 99.5% percentile of the loss distribution (as measured in USD millions). From Equation (7.1), this is the value of x that solves:

$$0.005 = 10{,}000x^{-3}$$

or[21]

$$\ln(0.005) = \ln(10{,}000) - 3\ln(x)$$

so that $\ln(x) = 4.836$ and $x = e^{4.836}$ (or about 126).

7.5 REDUCING OPERATIONAL RISK

Beyond measuring operational risk and determining appropriate capital levels, operational risk units should also try to be proactive in reducing both the probability of large losses and the severity of those losses when they occur. Financial institutions can learn from each other in this area. When a large loss occurs at one financial institution, risk managers throughout the world study what happened and consider the steps that can be taken to avoid a similar loss at their own organization.

Causes of Losses

Sometimes operational risk losses can be related to other factors that can be managed. For example, in some situations it might be possible to show that losses can be reduced by increasing employee training or the educational qualifications necessary for a certain position. In other situations, it might be possible to show losses arising from an outdated computer system.

[17] See B.V. Gnedenko, "Sur la distribution limité du terme d'une série aléatoire," Annals of Mathematics 44 (1943): 423–453.

[18] When many independent random variables are added it is well known that we get a normal distribution. Fat tails are likely to arise when a distribution is a result of many multiplicative effects. As an example of multiplicative effects, consider website usage. If people love a website, they are likely to recommend it to their friends. This creates a fat tail for the distribution of number of hits to a website in a day. Providing the fat tail is created by many people behaving independently in the way we have described, the power law will tend to hold.

[19] See P. De Fontnouvelle, V. DeJesus-Rueff, J. J. Jordan and E. S. Rosengren, "Capital and Risk: New Evidence on the Implications of Large Operational Risk Losses," Journal of Money, Credit and Banking, 38, 7 (October 2006): 1819–1846.

[20] Maximum likelihood methods and extreme value theory could be used to make these estimates from observed large losses. See P. Embrechts, C. Kluppelberg and T. Mikosch, *Modeling Extremal Events for Insurance and Finance*, New York: Springer, 1997.

[21] $\ln(ax^b) = \ln(a) + b\ln(x)$

As mentioned, it is not always the case that operational risk losses should be minimized. A cost-benefit analysis should be undertaken because the costs of reducing operational risk can sometimes outweigh the benefits. For example, a study might show that transaction processing errors can be reduced by 5% if a new computer system is developed in conjunction with extra training being given to employees. However, the cost of making the change might greatly exceed the present value of the reduction in losses.

Risk Control and Self Assessment

Risk control and self assessment (RCSA) is one way in which financial institutions try to understand operational risks while creating an awareness of operational risk among employees. The key term here is *self* assessment. Line managers and their staff, not operational risk professionals, are asked to identify risk exposures. The risks considered should include not just losses that have occurred in the past, but those that could occur in the future. There are many RCSA approaches:

- Interviewing line managers and their staff;
- Asking line managers to complete risk questionnaires;
- Reviewing risk incident history with line managers;
- Reviewing third-party reports such as those of auditors, regulators, and consultants;
- Reviewing reports of the experiences of similar managers in other companies;
- Using of suggestion boxes and intranet reporting portals;
- Implementation of a 'whistle blowing' process to encourage the reporting of risk issues; and
- Carrying out brainstorming in a workshop environment.

The assessment process should be repeated periodically (e.g., every year). The frequency of loss events and their severity should be quantified. Some loss events are an inevitable part of doing business. For others, the RCSA process may lead to improvements reducing the frequency of losses, the severity of losses, or both.

Key Risk Indicators

A developed understanding of the risks faced by line managers can lead to the development of key risk indicators (KRIs). These are data points that may indicate a heightened chance of operational risk losses in certain areas. In some cases, remedial action can be taken before it is too late. Simple examples of KRIs are metrics related to:

- Staff turnover,
- Failed transactions,
- Positions filled by temps, and
- Unfilled positions.

To use these indicators effectively, it is important to track how they change through time so that unusual behavior can be identified. Some KRIs are subtler than others. For example, the unwillingness of an employee to take vacations might be an indication that he or she could be engaged in unauthorized trading or embezzling funds. Some organizations have become quite sophisticated, using tools such as surveillance software to search for unusual email or phone activity indicative of an employee engaging in unlawful or unethical activity.

Education

Employee education can be important in reducing operational risk. We mentioned earlier how compliance is an area that can lead to huge operational risk losses. Educating employees about unacceptable business practices and (more importantly) creating a risk culture where such practices are perceived to be unacceptable is important. In 2007, Goldman Sachs received adverse publicity when it created a product (called ABACUS) that arguably benefited one client at the expense of another. The firm then took steps to change its risk culture and sent CEO, Lloyd Blankfein, on a 23-country tour of Goldman's regional offices. He spoke to employees about the importance of ethics and emphasized that the company should not sell products to clients unless they fully understand the range of possible outcomes.

Legal disputes are unfortunately an inevitable part of doing business. (This is particularly true in the litigious environment of the United States.) The in-house legal department within a financial institution needs to remind employees to be careful about what they write in e-mails and (when they are recorded) what they say in phone calls.[22] In a legal dispute where an organization is being sued, the organization usually must provide all relevant internal communications. Some can be very embarrassing. For example, Fabrice Tourre, who worked on the ABACUS product at Goldman Sachs, 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!!!"[23]

Before communicating via the use of emails or recorded phone calls, an employee should always consider whether he or she would be comfortable if the communication became public knowledge.

[22] Phone calls in a trading room are usually recorded.

[23] https://abcnews.go.com/GMA/Business/goldmans-blankfein-fabulous-fab-testify-today/story?id=10479173

7.6 INSURANCE

Many operational risks can be insured against. However, operational risk managers need to carefully evaluate whether the cost of insurance can be justified. As explained earlier, the new SMA for assessing operational risk is based on the frequency and magnitude of the losses incurred over the previous ten years. Thus, insuring against a loss can not only reduce the severity of losses, but also reduce capital requirements.

To understand how insurance companies view operational risk, we review the two key risks they face: moral hazard and adverse selection.

Moral Hazard

Moral hazard is the risk that the existence of an insurance contract will cause the insured entity to behave in a way that makes a loss more likely.

One example of moral hazard concerns rogue trader losses. If an insurance company insures a bank against such losses, it might be concerned traders would take large unauthorized risks. If a gain resulted, the bank would be pleased. If a loss resulted, a claim would be made against the insurance company.

In the light of this type of moral hazard, it is perhaps surprising that it is actually possible to buy insurance against rogue trader losses. In practice, insurance companies manage the moral hazard by carefully specifying how trading limits are implemented and monitored within banks. Rogue trader insurance policies are negotiated by the risk managers and insurance companies often require that the policies are not revealed to traders. Any losses incurred are investigated carefully, and if financial institutions fail to follow their insurance requirements, they might forfeit their payout.

More generally, insurance companies manage moral hazard by using deductibles so that a financial institution is responsible for the first part of any loss. There may also be a co-insurance provision where the insurance company pays only a percentage of a loss rather than the whole amount. Furthermore, there is always a limit on the total amount that can be paid out. Insurance premiums may also increase after a loss has been incurred.

Adverse Selection

Adverse selection is the problem an insurance company faces in distinguishing low-risk situations from high-risk situations. If it charges the same premium for a certain type of risk to all financial institutions, it will inevitably attract clients with the highest risk. Consider again the example of rogue trader insurance. If all banks were offered the same insurance premiums, banks with poor internal controls would tend to buy more insurance, while those with good internal controls would consider the cost of the insurance too high (and therefore buy less insurance).

Insurance against cyber risks provides another example of potential adverse selection. Those financial institutions with good cyber defenses are likely to consider cyber insurance to be too expensive, while those that have not invested heavily in this area will find the insurance attractive.

Insurance companies deal with adverse selection by finding out information about potential customers before providing a quote. Car insurance is a good example of this approach. A driver's initial price quote reflects past accidents, speeding tickets, etc. As time goes by, this information is updated, and the insurance premium is adjusted accordingly. In the case of rogue trader insurance and cyber insurance, a financial institution must convince an insurance company that it has good risk controls in place before it can qualify for insurance.

SUMMARY

Most risk managers and regulators consider operational risk to be the biggest risk faced by financial institutions. Bank regulators recognized this in 1999 and announced they would require a separate capital charge for operational risk. In 2016, they announced a revision to the way this capital charge is calculated. The new rules calculate capital based on a bank's ten-year operational risk loss history.

For economic capital purposes, potential operational risk losses are often calculated by first estimating loss frequency and loss severity, and then carrying out a Monte Carlo simulation. Scenario analysis is a useful tool for identifying the range of ways in which large losses can occur.

Information on operational risk losses can come from line managers and their staff. RCSA is a means in which this information can be collected in a systematic way. Sometimes, KRIs can be used to identify potential future operational risk losses.

Many operational risk losses can be insured against. Most policies include deductibles, coinsurance provisions, and policy limits. As a result, the policyholder is always left bearing part of the risk itself. Insurance premiums change through time as information on loss experience is collected.

A general point is that regulators have been successful in requiring financial institutions to devote more resources to managing operational risk. Operational risk groups at financial institutions have grown in size and importance and line managers have become more aware of the potential operational risk losses arising from their activities.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

QUESTIONS

Short Concept Questions

7.1 What is considered to be the largest operational risk facing financial institutions?

7.2 Give two examples of compliance risk at a bank that could lead to large fines.

7.3 What is the difference between the front office and the back office? Why should they be kept separate?

7.4 How did the basic indicator approach work in Basel II?

7.5 What were banks required to do in the AMA of Basel II?

7.6 Why did the Basel Committee abandon the AMA and replace it with the SMA?

7.7 What estimates are necessary to determine an operational loss probability distribution?

7.8 What is meant by moral hazard and adverse selection in insurance contracts?

7.9 Explain the power law.

7.10 What are two approaches can a financial institution use to reduce operational risk?

Practice Questions

7.11 How is moral hazard typically handled in insurance contracts?

7.12 If the average loss frequency is estimated as once every 24 months, what is the probability of (a) one loss in a year and (b) three losses in a year?

7.13 The 5 and 95 percentiles of a lognormal loss distribution are 50 and 200. What is the 99.7 percentile of the distribution?

7.14 The mean and standard deviation of a lognormal loss distribution are 100 and 40. What are the mean and variance of the logarithm of the loss?

7.15 The 95-percentile of a loss distribution is 20. Use the power law to obtain the 99-percentile if the α parameter is 5.

7.16 "Steps should always be taken to minimize operational risk." Discuss this statement.

7.17 A bank estimates from its own data that external fraud losses have (in USD millions) a mean of 50 and a standard deviation of 40. The data from a vendor shows that external fraud has a mean of 100 and a standard deviation of 800. It also shows that cyber risk has a mean of 300 and a standard deviation of 1,600. The bank has no data on cyber risk losses. How should it estimate the mean and standard deviation for its cyber risk losses?

7.18 In the last ten years, a bank has had losses (in millions of euros) of 4, 6, 12, 50, 70, 140, and 280. What is the bank's loss component under the SMA?

7.19 What is a rogue trader? What key risk indicator might be helpful in detecting a rogue trader?

7.20 What is risk control and self-assessment? List five ways it can be carried out.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

ANSWERS

Short Concept Questions

7.1 Cyber risk

7.2 Money laundering, terrorist financing, failure to comply with sanctions, assisting with tax evasion.

7.3 Trading takes place in the front office. Records are kept in the back office. If the two are not kept separate, traders might find ways of concealing losses.

7.4 Regulatory capital was 15% of the average gross income over the last three years. Gross income was defined as net interest income plus non-interest income.

7.5 Banks were required to set capital equal to the 99.9-percentile of the distribution of the loss from operational risk minus expected operational risk losses.

7.6 Different banks implemented the AMA in different ways. This meant that two different banks in identical positions could calculate quite different capital requirements.

7.7 Loss frequency and loss severity. Loss frequency can be described by the average number of losses per year. Loss severity can be described by the mean and standard deviation of losses when they occur.

7.8 Moral hazard is the risk that the existence of insurance will change the behavior of the insured party making claims more likely. Adverse selection is the risk that an insurance company is unable to distinguish good risks from bad.

7.9 The power law states that for a wide range of random variables, v, the probability that $v > x$ is approximately equal to Kx^{-a} for some constant parameters K, and α when x is large.

7.10 The use of key risk indicators, risk control and self assessment, and learning from the mistakes of others.

Solved Problems

7.11 Moral hazard is handled with deductibles, co-insurance provisions, policy limits, and tying premiums to past losses. In some cases, it may be necessary for the insurance company to require certain behavior from the policy holder if the policy is to remain valid. For example, in the case of rogue trader insurance there is likely to be a requirement concerning the way risk limits are enforced and a requirement that the existence of the policy is not communicated to the trading room.

7.12 In this case, the loss frequency λ is 0.5 per year. The probability of one loss in a year is

$$\frac{e^{-0.5} \times 0.5}{1!} = 0.303$$

The probability of three losses is

$$\frac{e^{-0.5} \times 0.5^3}{3!} = 0.0126$$

7.13 The 5 and 95-percentiles of the logarithm of the loss are ln(50) and ln(200), (i.e., 3.912 and 5.298). The logarithm of the loss has a mean of $(3.912 + 5.298)/2 = 4.605$ and a standard deviation of $(5.298 - 4.605)/1.645 = 0.421$. The 99.7-percentile of the logarithm of the loss is $4.605 + N^{-1}(0.997) \times 0.421 = 5.763$. The 99.7-percentile of the loss is therefore $e^{5.763} = 318.3$

7.14 In this case $w = 0.4^2 = 0.16$, so that the mean and variance of the logarithm of the loss are $\ln(100/\sqrt{1.16}) = 4.53$ and $\ln(1.16) = 0.148$.

7.15 We know that

$$0.05 = K \times 20^{-5}$$

so that $K = 160{,}000$. The 99-percentile is obtained by solving the following equation for x:

$$0.01 = 160000x^{-5}$$

It is 27.6.

7.16 Whether a reduction in operational risk is warranted should be determined by a cost-benefit analysis.

7.17 In the absence of any other data, the bank would assume that the mean cyber loss is 300/100 or three times its mean external fraud loss (i.e., 150), and that the standard deviation of cyber risk losses is 1,600/800, or twice its standard deviation of external fraud loss (i.e., 80).

7.18 In millions of euros, total losses for ten years are 562, losses greater than ten are 552, and losses greater than 100 are 420. The loss component is

$$7 \times 56.2 + 7 \times 55.2 + 5 \times 42.0 = 989.8$$

The following questions are intended to help candidates understand the material. They are *not* actual FRM exam questions.

7.19 A rogue trader is a trader who does unauthorized trading and finds a way of hiding losses. Failure to take any holidays could be an indication of a rogue trader.

7.20 RCSA involves determining operational risk from line managers and their staff. Approaches mentioned in the chapter are

- Interviewing line managers and their staff;
- Asking line managers to complete risk questionnaires;
- Review of the history of risk incidents with line managers;
- Reviewing with line managers third-party reports such as those of auditors, regulators, and consultants;
- Reviewing with line managers reports of the experiences of similar managers in other companies;
- The use of suggestion boxes and intranet reporting portals;
- Implementation of a 'whistle blowing' process to encourage the reporting of risk issues; and
- Carrying out brainstorming in a workshop environment.

8

Stress Testing

Learning Objectives

After completing this reading, you should be able to:

- Describe the rationale for the use of stress testing as a risk management tool.
- Describe the relationship between stress testing and other risk measures, particularly in enterprise-wide stress testing.
- Describe stressed VaR and stressed ES, including their advantages and disadvantages, and compare the process of determining stressed VaR and ES to that of traditional VaR and ES.
- Explain key considerations and challenges related to developing stress testing scenarios and building stress testing models.
- Describe reverse stress testing and describe an example of regulatory stress testing.
- Describe the responsibilities of the board of directors, senior management, and the internal audit function in stress testing governance.
- Describe the role of policies and procedures, validation, and independent review in stress testing governance.
- Describe the Basel stress testing principles for banks regarding the implementation of stress testing.

Stress testing is a risk management activity that has become increasingly important since the 2007–2008 financial crisis.[1] It involves evaluating the implications of extreme scenarios that are unlikely and yet plausible. A key question for a financial institution is whether it has enough capital and liquid assets to survive various scenarios. Some stress tests are carried out because they are required by regulators. Others are carried out by financial institutions as part of their internal risk management activities.

Earlier chapters have discussed how measures such as value-at-risk (VaR) and expected shortfall (ES) are calculated and used. Stress tests provide additional information for risk managers. When used in conjunction with VaR/ES analyses, they provide a more detailed picture of the risks facing a financial institution. Their advantage is that they can consider the impact of scenarios that are quite different from (and more severe than) the scenarios considered by VaR or ES. As explained in Chapter 2, bank regulators have moved toward basing market risk capital on stressed VaR and (more recently) stressed ES. These are measures based on how market variables behaved during a 12-month period that would have been significantly stressful for a firm's current portfolio.

This chapter discusses how stress-testing scenarios are generated internally by financial institutions. It also covers regulatory requirements, stress-testing principles published by the Basel Committee, and governance issues.

8.1 STRESS TESTING VERSUS VaR AND ES

VaR and ES are based on the estimated loss distribution. VaR allows a financial institution to reach a conclusion in the form of:

> We are X percent certain that our losses will not exceed the VaR level during time T.

In the case of ES, the conclusion is

> If our losses do exceed the VaR level during time T, the expected (i.e., average) loss will be the ES amount.

One disadvantage of VaR and ES is that they are usually backward-looking. They assume the future will (in some sense) be like the past. Stress testing, however, is designed to be forward-looking and answer more general "What If?" questions. Unlike VaR and ES, stress testing does not provide a probability distribution for losses. While it may be possible for management to assess probabilities for different scenarios, using stress tests to derive the full range of all possible outcomes is not usually possible.

Risk managers therefore have two types of analyses available to them. One is a backward-looking analysis where a loss distribution can be estimated. The other is a forward-looking analysis where different scenarios are assessed. The backward-looking VaR/ES analysis looks at a wide range of scenarios (some good for the organization and some bad) that reflect history. On the other hand, stress testing looks at a relatively small number of scenarios (all bad for the organization). There are other differences between the stress-testing approach and VaR/ES analyses. In the case of market risk, the VaR/ES approach often has a short time horizon (perhaps only one day), whereas stress testing usually looks at a much longer period.

The objective in stress testing should be to obtain an enterprise-wide view of the risks facing a financial institution. Often, the scenarios are defined in terms of macroeconomic variables such as GDP growth rates and unemployment rates. The impact of these variables on all parts of the organization must be considered along with the interactions between different areas. The overarching objective in stress testing is to determine whether the financial institution has enough capital and liquidity to survive adverse events.

Stressed VaR and Stressed ES

The distinction between stress testing and VaR/ES measures is blurred by measures known as stressed VaR and stressed ES. These have been mentioned in earlier chapters and will now be reviewed again.

VaR and ES have traditionally been calculated using data from the preceding one to five years. Daily movements in risk factors during this period are used to calculate potential future movements. In stressed VaR and stressed ES, however, this data is gathered from particularly stressful periods.[2] Stressed VaR and stressed ES therefore produce conditional loss distributions and

[1] A summary of the definitions of a scenario, a sensitivity, and a stress test provided by the International Actuarial Association is as follows. "A scenario projects one or more risk factors over multiple time periods for a firm or the economy. A sensitivity is the effect of alternative assumptions regarding a future environment. A stress test is the projection of the financial condition of a firm or the economy for an extremely adverse but plausible scenario." See http://www.actuaries.org/CTTEES_SOLV/Documents/StressTestingPaper.pdf

[2] The Basel Committee rules require banks to base stressed VaR and stressed ES on a 12-month period of significant stress for the bank's current portfolio.

conditional risk measures. Specifically, they are conditional on a repeat of a given stressed period and can be considered a form of historical stress testing.

Although stressed VaR/ES and stress testing have similar objectives, there are important differences between them. Suppose the year 2008 is used as the stressed period. Stressed VaR would reach the conclusion:

> If we had a repeat of 2008, we are X% certain that losses over a period of T days will not exceed the stressed VaR level.

Stressed ES would reach the conclusion:

> If losses over T days do excess the stressed VaR level, the expected (i.e., average) loss is the stressed ES.

Typically, T is a short period (i.e., one to ten days). In contrast, stress testing usually has a longer time horizon. It seeks to answer the questions such as, "If the next year is a repeat of 2008, how would our organization survive?" or "If next year were like 2008 but twice as bad, how would we survive?" It does not consider what would happen during the worst T days of 2008. Rather, it considers the impact of the whole of 2008 being repeated.

Traditional VaR measures are designed to quantify the full range of possible outcomes and can therefore be back-tested. For example, suppose we have a procedure for calculating one-day VaR with 99-percent confidence. We can test this measure by seeing how well the procedure would have worked in the past. If we find that losses would have only exceeded the calculated VaR around 1% of the time, we can have some confidence in our results. However, it is not possible to back-test stressed VaR or the output from stress testing in this way because these measures focus on extreme outcomes, which we do not expect to observe with any particular frequency.

8.2 CHOOSING SCENARIOS

The first step in choosing a stress-test scenario is to select a time horizon. While one-day or one-week scenarios are occasionally considered, scenarios lasting three months to two years are more common. The time horizon should be long enough for the full impact of the scenarios to be evaluated, and very long scenarios can be necessary in some situations. For example, a pension plan or insurance company concerned about longevity risk might consider stress tests stretching over several decades.

As we will describe later, some scenarios are determined by regulators. At this stage, we focus on those that are chosen internally. We will explain several ways in which these scenarios can be generated.

Historical Scenarios

Scenarios are sometimes based on historical data, and it is assumed that all relevant variables will behave as they did in the past. When discussing the historical simulation approach to determining VaR and ES, we explained that for some variables (e.g., equity prices and exchange rates) it is appropriate to create a scenario that assumes the proportional changes observed in the past are repeated. For other variables (e.g., interest rates and credit spreads), it is appropriate to assume the actual changes from the past are repeated. A similar point applies here. Actual changes from the stressed period will be assumed to recur for some variables, while proportional changes will be assumed for others.

There are many historical scenarios that might be of concern to risk managers. The 2007–2008 U.S. housing-related recession, which led to serious problems for many financial institutions, is an obvious one to use. Scenarios that could give rise to other adverse outcomes include the plunge in oil prices seen in the second half of 2014 and the flight to quality following Russia's default on its bonds in August 1998.

Sometimes, a moderately adverse scenario from the past is made more extreme by multiplying the movements in all risk factors by a certain amount. For example, we could take what happened during a certain loss-making six-month period in the past and double (or triple) the movements in all relevant variables. When magnified in this way, the scenario might become a much more serious problem for a financial institution. However, this approach assumes there is a simple linear relationship between the movements in risk factors. This is not necessarily the case, however, because correlations between risk factors tend to increase as economic conditions become more stressed.

Sometimes, historical scenarios are based on what happened to all market risk factors over one day or one week. For example, the impact of a day like October 19, 1987 (when the S&P 500 fell by 22.3 standard deviations) could be assessed. If this is considered too extreme, a scenario could be created from the days around January 8, 1988 (when the S&P 500 fell by 6.8 standard deviations). Other dates with large movements in equity prices are September 11, 2001 (during the 9/11 terrorist attacks) and September 15, 2008 (when Lehman Brothers declared bankruptcy). For a big one-day movement in interest rates, April 10, 1992 (when ten-year bond yields moved by 8.7 standard deviations) could be used.

These short-horizon stress tests can be supplements to stressed VaR and stressed ES calculations. While stressed VaR and stressed ES consider extreme movements during just one stressed period, short-horizon stress tests can pick big movements from many different stressed periods in the past.

Stress Key Variables

One approach to scenario building is to assume that a large change takes place in one or more key variables. Changes that might be considered include:

- A 200-basis point increase in all interest rates,
- A 100% increase in all volatilities,
- A 25% decline in equity prices,
- A 4% increase in the unemployment rate, and
- GDP declining by 2%.

Other changes could involve factors such as exchange rates, commodity prices, and default rates.

For market risks, a financial institution's internal systems will provide the impact of relatively small changes in the form of Greek letters such as delta, gamma, and vega (which are discussed in later chapters). In the case of stress testing, however, the changes are so large that these measures cannot be used. Furthermore, whereas Greek letters quantify the risks arising from changes to a single market variable over a short period of time, stress testing often involves the interaction of several market variables over much longer periods.

Ad Hoc Stress Tests

The stress tests we have described so far are likely to be carried out on a regular basis (e.g., every month) and the results can provide a financial institution with a good indication of the robustness of its financial structure. But it is important for firms to develop other scenarios reflecting current economic conditions, the particular exposures of the financial institution, and an up-to-date assessment of possible future adverse events. History never repeats itself exactly, and managerial judgement is necessary to either generate new scenarios or modify existing scenarios based on past data.

The decision by the U.K. government to hold a vote on whether to leave the European Union could have led to an ad hoc stress test risk for a financial institution with major business interests in the U.K. There was no historical precedent for the vote and so historical scenarios would not have captured the risks involved. Prior to the vote in June 2016, most people did not expect a "leave" decision despite the fact that it was a plausible outcome. A scenario where the U.K. population votes to leave the European Union would therefore have constituted a valid stress scenario for a financial institution prior to June 2016.

Other ad hoc stress tests could consider the impact of a change in government policy on a key issue affecting a financial institution or a Basel regulation that would require more capital to be raised in a short period of time. Adverse scenarios suggested by professional economists should be considered carefully. In 2005–2006, many economists suggested (correctly as it turned out) that the U.S. housing market was experiencing a bubble that sooner or later would burst. Even if the board and senior management at a financial institution did not agree with this assessment, it would have made sense to use it as a stress scenario.

The boards, senior management, and economics groups within financial institutions are in a good position to use their understanding of markets, world politics, and current global uncertainties to develop adverse scenarios. One way of developing the scenarios is for a committee of senior management to engage in brain-storming sessions. Research suggests committees consisting of three to five members with different backgrounds work best.[3] A key role of the committee should be to recommend actions that can be taken to mitigate unacceptable risks.

Using the Results

It is important that senior management recognizes the importance of stress testing and incorporates it into its decision making. We will discuss the role of the board and the governance of stress testing later. At this stage, we note that involving senior management in building scenarios makes it more likely that the stress testing will be taken seriously and used for decision-making.

It should be emphasized that the purpose of stress testing is not just to produce output answering "What if?" questions. Senior management and the board should carefully evaluate stress-test findings and decide whether some form of risk mitigation is necessary. There is a natural human tendency for a decision-maker to base decisions on what he or she considers to be the most likely outcome and to regard alternatives to be so unlikely that they are not worth considering.[4] Stress testing should be used by the board and senior management to ensure that this does not happen.

8.3 MODEL BUILDING

It should be possible to observe how most of the relevant risk factors behaved during the stressed period when building a scenario. The impact of the scenario on a firm's performance can

[3] See R. Clemens and R. Winkler, "Combining probability distributions from experts in a risk analysis," *Risk Analysis*, 19, 2 (April 1999): 187–203.

[4] This is known as anchoring and is one of the many cognitive biases that have been listed by psychologists.

then be assessed in a fairly direct way. However, it may be necessary to use judgement in determining the ease with which the firm could raise more capital or improve its liquidity.

Scenarios constructed by stressing key variables (and ad hoc scenarios) typically specify movements in only a few key risk factors or economic variables. To complete the scenarios, it is necessary to construct a model to determine how a range of other variables can be expected to behave. The variables specified in the scenario definition are sometimes referred to as core variables, whereas the other variables are referred to as peripheral variables.

One approach is to carry out an analysis (e.g., linear regression) relating the peripheral variables to the core variables. However, it is important to recognize that the focus is on the relationship between variables in stressed market conditions (rather than normal market conditions). Stressed periods from the past are therefore likely to be most useful in determining the relevant relationships.[5]

For credit risk losses, data provided by rating agencies can be useful. Table 6.1, for example, shows data on the annual percentage default rates for all rated companies between 1981 and 2018. This can be related to economic variables (e.g., the GDP growth rate and unemployment rate) to determine the overall default rates that can be expected in different scenarios. This can then be scaled up or down to estimate default rates for the different categories of loans on a financial institution's books. A similar analysis can be carried out for recovery rates so that loss rates can be determined.

For assessing market risk losses, the relevant peripheral variables are likely to be those whose movements can be related to changes in core risk factors (e.g., interest rates and equity prices). In areas such as investment banking, profitability is likely to be related to equity prices and key economic variables such as GDP growth.[6]

Knock-On Effects

Analysts should consider not only a scenario's immediate consequences, but also what are referred to as *knock-on effects*. A knock-on effect reflects the impact of how firms (particularly other financial institutions) respond to an adverse scenario. In responding to the adverse scenario, the companies often take actions exacerbating adverse conditions.

Consider a scenario that might have been constructed around a possible US housing price bubble in 2005–2006. It could have been assumed house prices would decline by 5–10%, which in turn would have increased the loss on a bank's mortgage portfolio. In fact, the scenario led to much more severe outcomes as outlined below:

- Some houses were worth less than their outstanding mortgage. Even though the owners could afford to service a mortgage, many chose to default. In effect, they exercised an option to sell the house back to the lender for the amount outstanding on the mortgage.[7] The house was then sold by the lender. This increased the supply of houses on the market, making the decline in housing prices greater than it otherwise would have been. This increased the losses on mortgages and securities created from mortgages.
- There was a flight to quality where all risky assets were perceived to be less attractive. As a result, equity prices and corporate bond prices declined sharply. The decline in corporate bond prices meant credit spreads increased.
- Banks were concerned about the creditworthiness of other banks and were reluctant to engage in interbank lending. This increased funding costs for banks.

8.4 REVERSE STRESS TESTING

Stress testing involves constructing scenarios and then evaluating their consequences. Reverse stress testing takes the opposite approach: It asks the question, "What combination of circumstances could lead to the failure of the financial institution?"

One reverse stress-testing approach involves the use of historical scenarios. Under this approach, a financial institution would look at a series of adverse scenarios from the past and determine how much worse each scenario would have to be for the financial institution to fail. For example, it might conclude that a recession three times worse than the one in 2007–2008 would lead to failure. As already mentioned, simply multiplying the changes observed in all risk factors during the 2007–2008 recession by the same amount is an approximation because there are likely to be non-linearities. Ideally, a firm would use a more sophisticated model incorporating the tendency for correlations to increase as market conditions become more stressed.

Analyzing all risk factors to find a plausible combination leading to firm failure is not usually feasible. One approach is to define a handful of key factors (e.g., GDP growth rate, unemployment

[5] See J. Kim and C. C. Finger, "A stress test to incorporate correlation breakdown," *Journal of Risk*, 2, 3 (Spring 2000): 5–19.

[6] For a further discussion of the models that can be used see T. Bellini, "Stress Testing and Risk Integration in Banks," Elsevier, 2016 and A. Siddique and I. Hasan, "Stress Testing," Risk Books, 2013.

[7] This was a viable strategy in states where the laws were such that a bank only had the right to take ownership of the house in the event of a default, so that other assets of the owner were not at risk.

rate, equity price movements, and interest rates changes) and construct a model relating all other relevant variables to them. It is then possible to search iteratively over all factor combinations to determine scenarios leading to failure.

Reverse stress testing can be an input to the work of a stress-testing committee. The committee is likely to discard some of the scenarios generated by reverse stress testing as totally implausible while flagging others for further investigation.

8.5 REGULATORY STRESS TESTING

Up to now, our discussion has centered around stress tests designed by financial institutions themselves. Regulators in many jurisdictions (including the United States, the United Kingdom, and the European Union) also require banks and insurance companies to carry out specified stress tests. In the United States, for example, the Federal Reserve carries out a stress test of all banks with consolidated assets of over USD 50 billion. This is referred to as the Comprehensive Capital Analysis and Review (CCAR). Banks are required to consider four scenarios:

1. Baseline,
2. Adverse,
3. Severely adverse, and
4. An internal scenario.

In early 2020, the Federal Reserve described the scenarios planned for 2020:[8]

> The stress tests include two hypothetical scenarios: baseline and severely adverse. The severely adverse scenario this year features a severe global recession in which the U.S. unemployment rate rises by 6.5 percentage points to 10 percent, and elevated stress in corporate debt markets and commercial real estate.
>
> Additionally, banks with large trading operations will be required to factor in a global market shock component as part of their scenarios. This year's shock features, among other things, heightened stress to trading book exposures to leveraged loans. Additionally, firms with substantial trading or processing operations will be required to incorporate a counterparty default scenario component. The chart below shows the components that apply to each firm.
>
> "This year's stress test will help us evaluate how large banks perform during a severe recession, and give us increased information on how leveraged loans and collateralized loan obligations may respond to a recession," Vice Chair for Supervision Randal K. Quarles said.
>
> The severely adverse scenario and baseline scenarios are not forecasts. The severely adverse scenario describes a hypothetical sets of events designed to assess the strength of banking organizations. Similarly, the baseline scenario is in line with average projections from surveys of economic forecasters. Each scenario includes 28 variables—such as gross domestic product, the unemployment rate, stock market prices, and interest rates—covering domestic and international economic activity.

Banks must submit a capital plan, documentation to justify the models they use, and the results of their stress tests. If they fail the stress test because their capital is insufficient, they are likely to be required to raise more capital and restrict the dividends they can pay until they have done so.

Banks with consolidated assets between USD 10 billion and USD 50 billion are subject to the Dodd-Frank Act Stress Test (DFAST). The scenarios in DFAST are like those in CCAR. However, banks are not required to submit a capital plan (as capital management is based on a standard set of assumptions).

By choosing the scenarios, bank regulators can evaluate the ability of different banks to survive adverse conditions in a consistent way. But they make it clear that they also want to see scenarios developed by the banks themselves that reflect their particular vulnerabilities.

8.6 GOVERNANCE

Governance is an important part of stress testing. The governance process should determine the extent of the stress testing carried out by a financial institution. It should also ensure that the assumptions underlying the tested scenarios have been carefully thought out, that the results are prudently considered by senior management, and that actions based on the results are taken when appropriate.

The Board and Senior Management

The governance structure within a financial institution is likely to depend on the legal, regulatory, and cultural norms within a country. Generally, there should be a separation of duties between the board of directors and senior management. The board of directors has the responsibility to oversee the key strategies. It is also responsible for the firm's risk appetite (i.e., the amount and type of risk an organization is willing to take to meet its strategic objectives) and risk culture (i.e., the financial institution's norms along with the collective attitudes and behaviors of its employees).

Stress testing is an important way in which risks are assessed within an organization. The board should define how stress testing is carried out. Specifically, it should determine the

[8] *Source:* the Federal Reserve.

procedures used to create the scenarios as well as the way in which assumptions and models are used to evaluate them. Board members do not carry out stress testing themselves, but they should be sufficiently knowledgeable to ask penetrating questions. They should feel free to use their own experience and judgement to ask for changes in the assumptions underlying the scenarios (or even to ask for totally new scenarios to be considered). When key decisions to mitigate risks are required, the board should feel free to ask for other analyses to supplement the stress-testing results.

Senior management is responsible for ensuring the stress-testing activities authorized by the board are carried out by competent employees as well as periodically reporting on those activities to the board. Senior management is also responsible for ensuring the organization is adhering to the appropriate policies and procedures.

It is tempting to use the same scenarios each time a stress test is carried out. However, senior management should ensure that the scenarios change as the economic environment changes and as new risks appear on the horizon. Stress testing should not be done mechanically just to satisfy the board and regulators. It should be an important part of the firm's decision-making and risk-mitigation strategies.

Senior management should have a deep understanding of how stress tests are carried out and should be in an even better position than the board to challenge key assumptions and models (or suggest new scenarios for consideration). Stress testing should not be done in a routine way using the same set of assumptions each time. Rather, the methods used should be refined over time. Even if the nature of a scenario is not changed, consideration should be given to changing its severity in light of changing circumstances. For example, a scenario where there is a 20% decline in equity prices might be changed to one where there is a 30% decline as volatilities increase.

It is important for the board and senior management to ensure stress testing covers all business lines and exposures. The same scenarios should be used across the whole financial institution, and the results should then be aggregated to provide an enterprise-wide view of the risks. Sometimes there will be offsets, but a scenario that leads to losses in one part of the business can do so in other parts as well. A range of different time horizons should be also considered because some adverse scenarios materialize more quickly than others.

Financial institutions must keep sufficient capital and liquid assets to survive stressful situations. Key outputs from a stress test are therefore the scenario's impact on capital and liquidity. Senior management and the board should carefully consider whether the results of stress tests indicate that more capital should be held or that liquidity should be improved. They should keep in mind that once an adverse scenario is underway, they are likely to have much less flexibility in managing capital and liquidity.

Policies and Procedures

A financial institution should have written policies and procedures for stress testing and ensure that they are adhered to. These policies and procedures should be clearly stated and comprehensive to ensure that different parts of the organization approach stress testing in the same way. The policies and procedures should

- Describe why stress testing is carried out,
- Explain stress-testing procedures to be followed throughout the company,
- Define the roles and responsibilities for those involved in stress testing,
- Define the frequency at which stress testing is to be performed,
- Explain the procedures to be used in building and selecting scenarios,
- Explain how independent reviews of the stress-testing function will be carried out,
- Provide clear documentation on stress testing to third parties (such as regulators, external auditors, and rating agencies) as appropriate,
- Indicate how the results of stress testing are to be used and by whom,
- Be updated as appropriate as stress-testing practices will change as market conditions change,
- Allow management to track how the results of stress tests change through time, and
- Document the operation of models and other software acquired from vendors or other third parties.

Documenting activities within a financial institution is often not a popular task. It is usually viewed as a less interesting and less creative activity than, say, building a model to investigate the impact of a recession where GDP declines. However, documentation is important because it ensures continuity if key employees leave and satisfies the needs of senior management, regulators, and other external parties.

Validation and Independent Review

Stress-testing governance should include independent review procedures. The reviews themselves should be unbiased and provide assurance to the board that stress testing is being carried out in accordance with the firm's policies and procedures. In this context, it is worth noting that a financial institution uses many different models and, whether they are part of the stress-testing procedures or not, they have to be subject to independent review to ensure that they are operating as intended.[9]

It is important that the reviewers of stress-testing procedures be independent of the employees conducting the stress test. The review should

- Cover the qualitative or judgemental aspects of a stress test,
- Ensure that tests are based on sound theory,
- Ensure that limitations and uncertainties are acknowledged, and
- Monitor results on an ongoing basis.

It is also important to ensure models acquired from vendors are subject to the same rigorous review as internal models.

The validation of stress-testing models is more difficult than the validation of other models because stress testing deals with rare events. As mentioned earlier, a VaR model with a one-day time horizon and 99-percent confidence level can be validated by counting the percentage of times actual losses would have exceeded the VaR level if the model had been used in the past. There is no similar way of validating the output from a stress test. Other validation approaches are also difficult because the limited amount of data available from previous stressed situations.

As we have explained, models describing the relationship between variables in normal market conditions may not describe how they behave in stressed market conditions. For example, correlations tend to increase in stressed market conditions and recovery rates tend to decline. The independent review should ensure that these phenomena are incorporated into stress-testing models.

The independent review should reach conclusions on the conceptual soundness of the stress-testing approach. When there are doubts about the best model to use in a certain situation, the results from several models can be compared and the totality of stress-testing reports can make the attendant uncertainties clear. It is often more appropriate to provide a range of possible losses rather than a single estimate. Expert judgement should be used to ensure results are presented in a way that is most useful for decision-making.

Internal Audit

The internal audit function has an important role to play in stress-testing governance. It should ensure that stress tests are carried out by employees with appropriate qualifications, that documentation is satisfactory, and that the models and procedures are independently validated. The internal audit function is not responsible for conducting the stress testing itself. Instead, it assesses the practices used across the whole financial institution to ensure they are consistent. Sometimes it will be able to find ways in which governance, controls, and responsibilities can be improved. It can then provide advice to senior management and the board on changes it considers to be desirable.

8.7 BASEL STRESS-TESTING PRINCIPLES

The publications of the Basel Committee have emphasized the importance of stress testing. The Basel Committee requires market risk calculations based on internal VaR and ES models to be accompanied by "rigorous and comprehensive" stress testing. Similarly, banks using the internal ratings-based approach in Basel II to determine credit risk capital are required to conduct stress tests to assess the robustness of their assumptions.

In May 2009, the Basel Committee published stress-testing principles for banks and their supervisors.[10] The principles were very much influenced by the 2007–2008 crisis, and they emphasize the importance of stress testing in determining how much capital is necessary to absorb losses from large shocks.

The principles note that stress testing plays an important role in:

- Providing forward-looking assessments of risk,
- Overcoming the limitations of models and historical data,
- Supporting internal and external communications,
- Feeding into capital and liquidity planning procedures,
- Informing and setting of risk tolerance, and
- Facilitating the development of risk mitigation or contingency plans across a range of stressed conditions.

Banks were arguably lulled into a false sense of confidence by the quiet economic conditions in the years preceding the crisis. It is therefore not surprising that the Basel Committee considers stress testing particularly important after long periods of benign conditions. The crisis showed that such conditions can lead to complacency and the underpricing of risk.

In examining the shortcomings of the stress testing carried out prior to the 2007–2008 crisis, the Basel Committee reached several conclusions. They can be summarized as follows.

[9] See, for example, 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.

[10] See Basel Committee on Banking Supervision, "Principles for Sound Stress-Testing Practices and Supervision," May 2009.

- The involvement of the board and senior management is important. 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. The Basel Committee notes that the banks that fared well in the financial crisis had a senior management that 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. At some banks, stress testing was a mechanical exercise that did not influence decision-making to any great extent and did not take account of changing business conditions. Sometimes, stress tests were carried out within business lines without interactions being considered and enterprise-wide results being produced. Banks should have the ability to aggregate exposures and respond quickly as problems emerge.
- The stress-testing methodologies used at some banks did not enable exposures in different parts of the bank to be aggregated. Experts from different parts of the banks did not cooperate in producing an enterprise-wide risk view. For example, it would have made sense for the optimism of the mortgage-backed securities traders to be tempered by retail lenders. The methodologies assumed average relationships between risk factors that had been observed in the past could be expected to continue to hold in the future. The knock-on effects that were mentioned earlier in this chapter were not considered.
- The scenarios chosen in the stress tests proved to be too mild and had durations that were too short. Additionally, they underestimated the correlations between different risk types, products, and markets. There was too much reliance on historical scenarios and not enough consideration of the risks created by introduction of new products and the new positions taken by the banks.
- Particular risks were not covered in sufficient detail in the scenarios. For example, risks relating to structured products, products awaiting securitization, imperfect hedging, and counterparty credit risk were not fully considered. The impact of a stressed scenario on liquidity was underestimated. The crisis gave rise to systemic risks; banks hoarded liquidity and were unwilling to advance loans to other banks in the way they would in normal market conditions.

After observing how stress testing had evolved since the crisis, the Basel Committee published a consultative document with a revised set of principles in 2017.[11] These may be used by national authorities to design stress-testing rules, guidance, or principles. The revised principles are summarized in the appendix and are consistent with the points we have made in this chapter.

SUMMARY

Stress testing is an important forward-looking risk management tool. Tools such as VaR and ES usually rely on the assumption that the future will be like the past and, as a result, may fail to consider key risks. Even stressed VaR and stressed ES are based on stressed periods that have actually occurred in the past. The scenarios considered by stress testing (if carried out properly) should include those that reflect the current business environment and those that have never happened before.

Senior managers are liable to argue that stress testing leads to risk assessments that are unduly pessimistic and capital requirements that are too high. However, risk management does not mean that no risks should be taken. It simply means that risks should be understood and that actions be taken when they are unacceptable. It is important that both the board and senior management be fully committed to stress testing and consider the results from stress testing when making decisions. Stress testing should not be a game between a financial institution and its regulators with the objective of producing less severe outcomes so that capital requirements are not increased.

There are many ways in which stress-testing scenarios can be developed. Some are based on stressed periods from the past and can be regarded as extensions of stressed VaR and stressed ES measures. Others consider the impact of big changes in key variables. The most important stress tests are those which are based on the judgement of senior management and consider the risks they see on the horizon. Some of these risks might be totally new and require the development of scenario-specific models.

[11] Basel Committee on Banking Supervision, "Stress Testing Principles," Consultative Document, December 2017.

APPENDIX

Basel Committee Stress Testing Principles[12]

1. ***Stress testing frameworks should have clearly articulated and formally adopted objectives***

   Stress testing frameworks should be designed to meet clear objectives that are documented and approved at the board level of the organization, or an appropriately senior-level governance body. The objectives should be the basis to set out requirements and expectations of the framework and should be consistent with the risk management framework of the bank and its overall governance structure.

   Staff that are involved in the implementation of stress testing frameworks should also have a clear understanding of the objectives of the framework, as this will help to guide any discretionary or judgemental elements.

2. ***Stress testing frameworks should include an effective governance structure***

   Stress testing frameworks should include an effective governance structure that is clear, comprehensive and documented. This should specify the roles and responsibilities of senior management, oversight bodies, and those responsible for the ongoing operation of the stress testing framework. This governance framework should ensure a full and consistent oversight and monitoring of the actions taken at the different stages of the stress testing process.

   Roles and responsibilities should be specified for all aspects of the stress testing framework, including: scenario development and approval, model development and validation, reporting and challenge of results, and the use of stress test outputs. The roles of the second and third lines of defence should be specified (i.e., risk management and compliance, and internal audit, respectively). Policies and procedures should cover all aspects of the stress testing framework, be clearly documented, kept up-to-date and be approved by the board and/or senior management.

   The stress testing framework should also ensure collaboration of all necessary stakeholders and the appropriate communication to stakeholders of the stress testing assumptions, methodologies, scenarios and results. The engagement structure should facilitate credible challenge of the stress testing framework, both at senior and technical expert levels, including not only assumptions, methodologies, scenarios and results, but also the assessment of its ongoing performance and effectiveness, and the remediation of gaps identified by key stakeholders.

3. ***Stress testing should be used as a risk management tool and to inform business decisions***

   Stress testing is a forward-looking risk management tool that constitutes a key input into banks' and authorities' activities related to risk identification, monitoring, and assessment. As such, stress testing should also contribute to formulating and pursuing strategic and policy objectives.

   When using the results of stress tests, banks and authorities should have a clear understanding of their key assumptions and limitations, for instance in terms of scenario relevance, risk coverage, and model risk.

   To be a meaningful risk management tool, stress tests should be undertaken regularly. While ad hoc stress tests may be performed for specific reasons, generally stress tests should be undertaken, according to a well-defined schedule. The appropriate frequency will depend on several factors, including: the objectives of the stress test framework, the scope of the stress test, the size and complexity of the bank or banking sector, as well as changes in the macroeconomic environment.

4. ***Stress testing frameworks should capture material and relevant risks and apply stresses that are sufficiently severe***

   Stress testing frameworks should capture material and relevant risks, as determined by a sound risk identification process. The risk identification process should include a comprehensive assessment of risks, including those deriving from both on- and off-balance sheet exposures, earnings vulnerabilities, and other factors that could affect the solvency or liquidity position of the bank (or banks in the case of supervisory stress tests).

   Stress test scenarios should be designed to capture material and relevant risks identified in the risk identification process and key variables within each scenario should be internally consistent. A narrative should articulate how the scenario captures the risks. If certain material and relevant risks are excluded from the scenarios, their exclusion should be explained and documented. The scenarios should be

[12] This appendix is a set of extracts from Basel Committee on Banking Supervision, "Stress Testing Principles," Consultative Document, December 2017. Used by permission of Bank for International Settlements.

sufficiently severe and varied, given the objective of the exercise, to provide a meaningful test of the resilience of banks. That is, the scenarios should be sufficiently severe but plausible.

The scenarios and sensitivities that are used in stress tests should be reviewed periodically to ensure that they remain relevant. Consideration should be given to historical events and hypothetical future events that take into account new information and emerging risks in the present and foreseeable future. 'Ahistorical' scenarios may be warranted if new or heightened vulnerabilities are being identified, or if historical data do not contain a severe crisis episode. The scenarios and the sensitivities should also take into account the current macroeconomic and financial environment.

5. ***Resources and organisational structures should be adequate to meet the objectives of the stress testing framework***

   Stress testing frameworks should have organisational structures that are adequate to meet their objectives. Governance processes should ensure the adequacy of resourcing for stress testing, including ensuring that resources have the appropriate skill sets to execute the framework. Resourcing decisions should take account of the fact that stress tests have become more sophisticated over time, increasing the need for specialised staff, systems, and IT infrastructure.

   Processes to ensure resources have the appropriate skill sets could include building the skills of internal staff, knowledge transfer to internal staff, as well as hiring personnel with specialised stress testing skills. The set of skills typically required include (but are not limited to) expertise in: liquidity risk, credit risk, market risk, capital rules, financial accounting, modelling, and project management.

6. ***Stress tests should be supported by accurate and sufficiently granular data and by robust IT systems***

   In order for risks to be identified and the results of stress tests to be reliable, the data used should be accurate and complete, and available at a sufficiently granular level and in a timely manner.

   Both banks and authorities should have in place a robust *data* infrastructure capable of retrieving, processing, and reporting information used in stress tests that ensure that the information is of adequate quality to meet the objectives of the stress testing framework. Processes should be in place to address any identified material information deficiencies.

7. ***Models and methodologies to assess the impacts of scenarios and sensitivities should be fit for purpose***

   The models and methodologies used to derive stress estimates and impacts should be appropriate for the purpose and intended use of the stress tests. This implies:

   - the need to adequately define at the modelling stage the coverage, segmentation and granularity of the data and types of risks in line with the objectives of the stress test framework;
   - the level of sophistication of the models should be appropriate for both the objectives of the exercise and the type and materiality of the portfolios being monitored using the models; and
   - the models and other methodologies used for stress tests should be well-justified and documented.

   Sound model development requires the collaboration of different experts. The model developers should engage with stakeholders to gain insights into the risks being modelled and to identify the business objectives, business drivers, risk factors and other associated business information that are relevant given the objectives of the stress testing framework (e.g., market, product or portfolio types, nature and materiality of risk exposures). The modelling choices and calibration decisions should consider the interactions between different risk types, as well as the linkages between models. In this regard, the links between solvency and liquidity stresses should be considered. The collaboration of model developers and stakeholders is particularly important for bank-wide stress testing to ensure the inclusion of all material risks and a sound aggregation of results.

   Stress tests employ a certain amount of expert judgement, including assumptions within a model or methodology. In some cases, model overlays are appropriate. Like the models, these overlays or expert judgements should be well-justified, documented and subject to credible challenge.

8. ***Stress testing models, results and frameworks should be subject to challenge and regular review***

   Regular review and challenge are key steps in the stress testing process for both banks and authorities. They are critical to improving the reliability of stress test results, aiding an understanding of their limitations, identifying areas where the stress testing approach should be improved and ensuring that the stress test results are being used in a way that is consistent with the framework's objectives. Such reviews should provide coverage of all aspects of the stress testing framework on a periodic basis and should be used to ensure that stress testing frameworks are maintained and regularly updated.

9. ***Stress testing practices and findings should be communicated within and across jurisdictions***

   Communication of stress testing activities across relevant internal and external stakeholders can have benefits for both banks and supervisors. Sharing of results can, where appropriate, provide important perspectives on risks that would not otherwise be available to an individual entity or authority.

   Disclosure of results of stress tests, whether by banks or authorities, can help to improve market discipline and provide confidence in the resilience of the banking sector to identified stresses. Banks and authorities that choose to disclosure stress test results should carefully consider ways to ensure that market participants understand data that is disclosed, including limitations and assumptions on which it is based. This will help to reduce the risk that market participants draw ill-informed conclusions about the resilience of banks with differing or negative results.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

QUESTIONS

Short Concept Questions

8.1 Explain the information provided by VaR, stressed VaR, and stress testing.

8.2 What are three different ways scenarios can be generated for stress testing?

8.3 What is reverse stress testing?

8.4 Explain knock-on effects in stress testing.

8.5 Explain the difference between a peripheral variable and a core variable in stress testing.

8.6 Which banks does CCAR apply to?

8.7 What is DFAST short for? Which banks does it apply to?

8.8 Is a capital plan required as part of (a) CCAR and (b) DFAST?

8.9 What are the three types of scenarios considered by CCAR?

8.10 Distinguish briefly the role of the board and senior management in a well-designed stress-testing framework.

Practice Questions

8.11 In what ways was the stress testing carried out prior to the 2007–2008 crisis inadequate?

8.12 List five things that should be included in the policies and procedures for stress testing.

8.13 Explain the role of (a) validation and independent review and (b) an internal audit in stress testing.

8.14 "It is important that risk managers consider the same scenarios each month so that trends can be identified." Discuss this statement.

8.15 What is the difference between stressed VaR, stressed ES, and stress testing?

8.16 A mildly adverse scenario can be made more severe by magnifying the movements in all relevant variables. What are the potential pitfalls in doing this?

8.17 What are the consequences to a bank of failing a regulatory stress test, such as CCAR?

8.18 To ensure its survival, a financial institution should focus on two key outputs from stress testing. What are they?

8.19 Why do you think some shareholders want the chairman of the board of a company to be a different person from the chief risk officer?

8.20 How might loan losses be determined from a scenario where GDP growth and unemployment are specified?

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

ANSWERS

Short Concept Questions

8.1 VaR provides a high percentile of the distribution of losses over a short period based on recent history. Stressed VaR provides a high percentile of the distribution of losses over a short period conditional on a repeat of a stressed period. Stress testing evaluates the outcome from a particular stress scenario over a longer period.

8.2 Historical data, stressing key variables, and developing ad hoc scenarios reflecting the current business environment.

8.3 Reverse stress testing searches for ways in which an organization can fail.

8.4 Knock-on effects are secondary implications of an adverse scenario caused by the way businesses respond to it.

8.5 Core variables are those for which explicit forecasts are made. Peripheral variables are other variables whose behavior must be deduced from the behavior of core variables.

8.6 CCAR applies to banks with over USD 50 billion of consolidated assets.

8.7 DFAST stands for Dodd-Frank Act Stress Test. It applies to banks with consolidated assets between USD 10 billion and USD 50 billion.

8.8 CCAR requires a capital plan; DFAST does not.

8.9 The three scenarios considered by CCAR are baseline, adverse, and severely adverse.

8.10 The board should define how stress testing is to be carried out within a financial institution. Senior management is responsible for ensuring that the stress-testing activities authorized by the board are carried out.

Solved Problems

8.11 The Basel Committee observes that in some cases the board and senior management were not sufficiently involved, stress testing was a mechanical exercise, exposures were not aggregated to produce an enterprise-wide view, knock-on effects were not considered, stress tests were too mild and not long enough in duration, correlations were under-estimated, and there was too much reliance on historical data and not enough consideration of new products and new positions being taken.

8.12 The list given in the chapter is

- Describe why stress testing is carried out;
- Explain stress-testing procedures to be followed throughout the company;
- Define the roles and responsibilities;
- Define the frequency with which stress testing is to be performed;
- Explain the procedures to be used in defining the selection of scenarios;
- Explain how independent reviews of the stress-testing function will be carried out;
- Provide clear documentation on stress testing to third parties such as regulators, external auditors, and rating agencies as appropriate;
- Indicate how the results of stress testing are to be used and by whom;
- Be updated as appropriate because it is recognized that as market conditions change, stress-testing practices will also change;
- Allow management to track how the results of stress tests change through time; and
- Document the operation of models and other software acquired from vendors or other third parties.

8.13 The validation and internal review process should carefully evaluate the models and procedures to determine their soundness and that the limitations are acknowledged. The internal audit has a more "big picture" role. It should ensure that stress tests are carried out by well-qualified people and that procedures are consistent across the organization. It looks for ways in which governance, controls, and responsibilities can be improved and provides advice to both senior management and the board.

8.14 This is not true. Scenarios should not be chosen mechanically. They should be adjusted as the business environment changes.

8.15 Stressed VaR calculates a percentile of the distribution of losses over a short period of time conditional on a stressed scenario from the past recurring. Stressed ES is the average loss conditional on the loss being greater than the stressed VaR level in the stressed scenario. Stress testing looks at the full consequences of a particular stress scenario that may or may not have occurred in the past. The time horizon is usually much longer than for stressed VaR/ES.

8.16 The relationship between variables may be complex. When the change in variable X observed in a certain scenario doubles, it may not be the case that the change in variable Y does so as well. Correlations increase in stressed market conditions.

8.17 It may have to raise more capital and be subject to some restrictions on the dividends it can pay.

8.18 It should focus on capital and liquidity.

8.19 The board makes key strategic decisions on what should be done and how it should be done. Senior management is responsible for implementing the decisions and reporting back. This is true in stress testing and other areas. If the same person leads senior management and the board, the separation of the responsibilities is not as clear-cut.

8.20 Default rate data year-by-year is provided by rating agencies. This can be related to data on GDP growth and unemployment to assess how adverse scenarios will affect the percentage of a bank's loans that default. Recovery rates can be similarly analyzed.

9

Pricing Conventions, Discounting, and Arbitrage

Learning Objectives

After completing this reading, you should be able to:

- Define discount factor and calculate present and future values using a discount function.
- Define the "law of one price," explain it using an arbitrage argument, and describe how it can be applied to bond pricing.
- Identify arbitrage opportunities for fixed-income securities with certain cash flows.
- Identify the components of a US Treasury coupon bond and compare the structure to Treasury STRIPS, including the difference between P-STRIPS and C-STRIPS.
- Construct a replicating portfolio using multiple fixed-income securities to match the cash flows of a given fixed-income security.
- Differentiate between "clean" and "dirty" bond pricing and explain the implications of accrued interest with respect to bond pricing.
- Describe the common day-count conventions used to calculate interest on a fixed-income security.

This is the first of several chapters where we discuss the fixed-income markets. These chapters extend and complement the chapter on interest rates in Financial Markets and Products.

Discount factors are the focus of this chapter. These are numbers that allow us to relate a cash flow received in the future to its value today. Suppose the price of a security indicates a cash flow of Y at a future time T is equivalent to a cash flow X today. The discount factor applicable for a maturity T is the factor we multiply Y by to get X (i.e., X/Y).

We show how discount factors can be extracted from Treasury bills and Treasury bonds. We also explain the law of one price and how it leads to arbitrage arguments. We introduce Treasury strips, which are zero-coupon instruments created from Treasury bonds. We describe how Treasury bill prices are quoted in the United States and discuss some of the day-count conventions used in the fixed-income markets. We also distinguish between "clean" and "dirty" bond prices.

Our discussion centers around the instruments traded in the United States. Similar instruments trade in many other countries. However, the conventions for pricing and quoting instruments vary from country to country.

9.1 TREASURY BILLS

Treasury bills are instruments issued by a government to finance its short-term funding needs. They last one year or less and are defined by:

- Their face value (also referred to as the principal amount or par value), and
- Their maturity date.

The holder of a Treasury bill receives the face value on the maturity date. Table 9.1 shows quotes for U.S. Treasury bills on March 9, 2018. The bid quote gives the price at which a market maker is prepared to buy the Treasury bill. The ask quote (sometimes referred to as the offer quote) gives the price at which a market maker is prepared to sell the Treasury bill.

As we shall see shortly, the quote for a Treasury bill is a measure of the interest rate earned. Treasury bill interest rates in the United States were considered to be quite low (compared with those in the past) in March 2018. But they declined very quickly to an extremely low level over the following two years.[1] By April, 2020 Treasury rates were only about 0.13% per annum. We will use data from March 9, 2018 in this chapter, rather than more recent data, because very low interest rates do not provide a good illustration of key points we want to make.

Table 9.1 **Treasury Bill Quotes on March 9, 2018**

Maturity	Bid	Ask
March 15, 2018	1.345	1.335
March 22, 2018	1.395	1.385
March 29, 2018	1.495	1.485
April 5, 2018	1.54	1.53
April 12, 2018	1.568	1.558
April 19, 2018	1.56	1.55
April 26, 2018	1.548	1.538
May 3, 2018	1.565	1.555
May 10, 2018	1.603	1.593
May 17, 2018	1.623	1.613
May 24, 2018	1.628	1.618
May 31, 2018	1.63	1.62
June 7, 2018	1.64	1.63
June 14, 2018	1.648	1.638
June 21, 2018	1.68	1.67
June 28, 2018	1.678	1.668
July 5, 2018	1.72	1.71
July 12, 2018	1.73	1.72
July 19, 2018	1.74	1.73
July 26, 2018	1.775	1.765
Aug. 2, 2018	1.803	1.793
Aug. 9, 2018	1.815	1.805
Aug. 16, 2018	1.823	1.813
Aug. 23, 2018	1.823	1.813
Aug. 30, 2018	1.828	1.818
Sept. 6, 2018	1.835	1.825
Sept. 13, 2018	1.813	1.803
Oct. 11, 2018	1.795	1.785
Nov. 8, 2018	1.813	1.803
Dec. 6, 2018	1.823	1.813
Jan. 3, 2019	1.9	1.89
Jan. 31, 2019	1.95	1.94
Feb. 28, 2019	1.97	1.96

Define Q as the quoted price of a Treasury bill and C as the cash price. The latter is the price that would apply to a Treasury bill with a face value of USD 100. The relationship between C and Q is

$$Q = \frac{360}{n}(100 - C)$$

[1] This was a result of policies followed by the U.S Federal Reserve. The U.S. Federal Reserve intervenes in the market to raise or lower rates. It chose to lower rates between March 2018 and March 2020 to stimulate the economy and deal with the pandemic that surfaced in early 2020.

This can be rewritten as:

$$C = 100 - \frac{Qn}{360} \tag{9.1}$$

where n is the number of calendar days until the maturity of the Treasury bill. To understand these formulas, we first note that Q is a measure of the interest earned on the Treasury bill. If $n = 360$ (so that the Treasury bill has 360 days to maturity), the cash price is $100 - Q$. The buyer would pay $100 - Q$ for the Treasury bill and receive 100 in 360 days. The quote Q is the interest earned over a 360-day period as a percentage of the face value. When $n < 360$, the interest earned over n days is scaled down so that it is $Qn/360$.

There are two somewhat unusual aspects of this. The first is that Q measures the interest over a 360-day period even though most investors would choose to calculate the interest earned over a whole year. The second is that the interest rate is calculated as a percentage of the face value even though most investors would choose to calculate interest earned as a percentage of the amount paid for the Treasury bill. The interest earned over 360 days as a percentage of the amount paid is

$$\frac{Q}{100 - Q}$$

More generally, the interest earned over n days as a percentage of the amount paid is

$$\frac{Qn/360}{100 - Qn/360}$$

These interest rates can be adjusted so that they become interest rates per 365 days (instead of interest rates per 360 days) by multiplying them by 365/360.

This is an example of the quotation conventions used in the interest rate market. These quotation conventions were determined many years ago when traders had neither computers nor hand calculators. Thus, it was easiest to do calculations based on a 360-day period and to calculate interest as a percentage of the face value.

To illustrate Treasury bill calculations, consider the June 7, 2018 Treasury bill in Table 9.1. The bid quote is 1.640 and the ask quote is 1.630. There are 90 days between March 9, 2018 and June 7, 2018. From Equation (9.1), the bid cash price is

$$100 - \frac{1.640 \times 90}{360} = 99.5900$$

The ask cash price is

$$100 - \frac{1.630 \times 90}{360} = 99.5925$$

This indicates that the investor can sell the Treasury bill to the market maker for USD 99.5900 per USD 100 of face value and buy the Treasury bill from the market maker for USD 99.5925 per USD 100 of face value. Note that in the first case, the market maker is buying at his or her bid price, and in the second case the market maker is selling at his or her ask price. The mid-market price is the average of the bid and ask prices. In this case, it is USD 99.59125 ($= 0.5 \times (99.5900 + 99.5925)$). Note that while the bid quote is greater than the ask quote, the bid cash price is less than the ask cash price.

The Treasury bill we have just considered gives us a way of translating the value of cash received in 90 days to its value today. For example, USD 1 million received on June 7, 2018 is worth USD 995,912.50 on March 9, 2018. Another way of saying this is that the 90-day discount factor on March 9, 2018 was 0.9959125. Risk-free cash flows to be received in 90 days should be multiplied by this amount to calculate their present value.

Calculating the future value of a cash flow is the opposite to discounting the cash flow. If USD 1,000 is owned on March 9, 2018, the future value (based on the mid-market discount rate) is

$$\frac{1000}{0.9959125} = 1004.10$$

9.2 TREASURY BONDS

We now move on to consider U.S. Treasury bonds. While a Treasury bill lasts less than one year from the time it is issued, a Treasury bond lasts more than one year. Bonds with a maturity between one and ten years are sometimes referred to as *Treasury notes*, but to keep the terminology simple we will refer to all coupon-bearing Treasury instruments as Treasury bonds. U.S. Treasury bonds are defined by:

- The face value (also referred to as the principal amount or par value),
- The coupon rate, and
- The maturity date.

The face value is the amount the bond holder receives on the maturity date. Coupon payments are usually made at half the coupon rate every six months. For example, if the coupon rate is 3%, USD 15 is paid per USD 1,000 of face value every six months. The coupon payment dates are the maturity date, six months before the maturity date, 12 months before the maturity date, 18 months before the maturity date, and so on. The maturity dates for Treasury bonds in the United States are usually the 15th of the month or the end of the month.

Table 9.2 shows quotes for some of Treasury bonds on March 9, 2018. The quotes are per USD 100 of face value. The bonds last between one week and 30 years. We can assume the Treasury bonds in Table 9.2 lasting less than one year on March 9, 2018 were issued by the U.S. government some time ago with maturities exceeding one year. They are now close to maturity.

Table 9.2 Treasury Bond Quotes on March 9, 2018

Maturity	Coupon	Bid	Ask
March 15, 2018	1	99.9922	100.0078
April 15, 2018	0.75	99.9141	99.9297
May 15, 2018	1	99.8906	99.9063
May 31, 2018	0.875	99.8125	99.8281
May 31, 2018	1	99.8359	99.8516
May 31, 2018	2.375	100.1406	100.1563
Aug. 15, 2018	4	100.8906	100.9063
Sept. 15, 2018	1	99.5313	99.5469
Oct. 15, 2018	0.875	99.3594	99.3750
Nov. 15, 2018	1.25	99.5000	99.5156
Dec. 15, 2018	1.25	99.4219	99.4375
Feb. 15, 2019	2.75	100.6094	100.6250
Aug. 15, 2019	0.75	97.9844	98.0000
Feb. 15, 2020	3.625	102.5000	102.5156
Feb. 15, 2020	8.5	111.8906	111.9063
Jan. 15, 2021	2	98.8672	98.8828
Jan. 31, 2022	1.5	96.0703	96.0859
Jan. 31, 2023	1.75	95.8750	95.8906
Feb. 15, 2024	2.75	100.0703	100.0859
Jan. 31, 2025	2.5	98.0938	98.1094
Feb. 15, 2026	1.625	91.2500	91.2656
Feb. 15, 2027	2.25	94.9297	94.9453
Feb. 15, 2028	2.75	98.7813	98.7969
Feb. 15, 2029	5.25	121.7422	121.8047
May 15, 2030	6.25	133.8828	133.9453
Feb. 15, 2031	5.375	126.1953	126.2578
Feb. 15, 2036	4.5	120.7813	120.8438
Feb. 15, 2037	4.75	124.9844	125.0469
Feb. 15, 2038	4.375	119.9219	119.9844
Feb. 15, 2039	3.5	106.7422	106.8047
Feb. 15, 2040	4.625	124.5391	124.6016
Feb. 15, 2041	4.75	126.9688	127.0313
Feb. 15, 2042	3.125	100.0938	100.1250
Feb. 15, 2043	3.125	99.7734	99.8047
Feb. 15, 2044	3.625	108.4141	108.4453
Feb. 15, 2045	2.5	87.9922	88.0234
Feb. 15, 2046	2.5	87.6719	87.7031

As with Treasury bills, it is important to distinguish between the cash price that is paid by the purchaser for a Treasury bond and the quoted price. In the case of Treasury bonds, the cash price is the quoted price plus accrued interest. The accrued interest is the interest earned between the most recent coupon date and the settlement date. It is calculated as:

$$c\frac{t}{T} \tag{9.2}$$

where c is the coupon to be paid on the next coupon date (equal to half the coupon rate applied to the face value), T is the number of calendar days between the last coupon date and the next coupon date, and t is the number of calendar days between the last coupon date and the settlement date.

Consider the bond maturing on February 15, 2028 in Table 9.2. The coupon is 2.75%. Coupons equal to 1.375%(= 0.5 × 2.75%) are paid on February 15 and August 15, each year. All transactions in Treasury bonds are settled after one business day (this is referred to as "T + 1" settlement). If the bond is purchased on Friday, March 9, 2018, it will be settled on Monday, March 12, 2018. The number of calendar days between February 15, 2018 and March 12, 2018 is 25. The number of calendar days between February 15, 2018 and August 15, 2018 is 181. This means that $t = 25$ and $T = 181$. The accrued interest per USD 100 of face value is therefore:

$$1.375 \times \frac{25}{181} = 0.1899$$

The bond quote is bid 98.7813 and ask 98.7969. When the accrued interest is added, the bid price becomes

$$98.7813 + 0.1899 = 98.9712$$

and the ask price becomes

$$98.7969 + 0.1899 = 98.9868$$

This indicates a market maker is prepared to buy the bond for USD 98.9712 and sell the bond for 98.9868. To put this another way, an investor could sell the bond to a market maker for USD 98.9712 and buy it from the market maker for USD 98.9868.

The quoted price of the bond is referred to as the *clean price* and the cash price paid for the bond is referred to as the *dirty price*. The dirty price reflects the fact the owner of a bond has earned interest that has not yet been paid. In our example, the owner of the bond has earned 25/181 of the next coupon because 25 days of the 181 days between coupon payment dates will have passed by the settlement date.

Traders like to use the quoted (clean) price rather than the cash (dirty) price for quotations because it is less variable. The clean price reflects only interest rate movements, whereas the dirty price reflects both interest rate movements and the build-up of accrued interest between coupon payment dates.

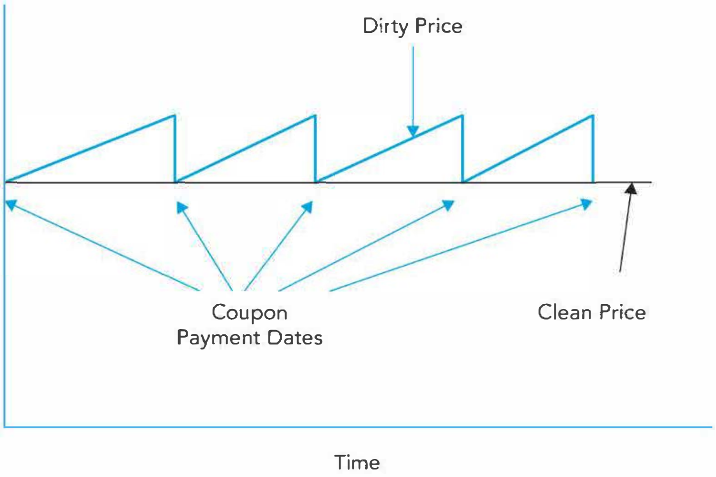


Figure 9.1 Changes in clean and dirty bond prices through time.

Consider a simple situation where the interest rate on Treasury securities is 3% for all maturities. A bond with a 3% coupon would be sold for its par value. If 3% continues to be the rate for all maturities, the bond will continue to sell for its par value. However, the dirty price will exhibit the pattern shown in Figure 9.1. Immediately after a coupon payment, the clean price equals the quoted price because there is no accrued interest. As time passes from one coupon payment date to the next, accrued interest is accumulated and the dirty price increases. Then, as soon as the next coupon is paid, it drops down.

9.3 SHORT POSITIONS

Prior to discussing the law of one price and arbitrage, it is appropriate to explain short positions. Suppose an investor instructs a broker to short 100 shares of a stock currently worth USD 50 per share. The broker will borrow the shares from an investor who owns them and sell them in the market. At some stage, the investor must buy the shares that have been borrowed so the broker can replace them in the account of the person from whom they have been borrowed. Often, a small fee is charged when shares or other assets are borrowed for shorting.

The investor who has shorted must pay any income (such as dividends or interest) on the shorted securities. This income is transferred to the person from whom the securities were borrowed. Consider again the investor that has shorted 100 shares of a stock when the price is USD 50 per share. Suppose two months later, the price has dropped to USD 40 per share and a dividend of USD 2 per share was paid after one month. If the investor closes out the short position at the end of the two months, the investor gains USD 8 per share, or USD 800 in total. This is because 100 shares were sold for USD 50 per share, they were repurchased for USD 40 per share, and USD 2 per share was paid in dividends (assuming no fee for borrowing the shares). The investor has gained because the price of the shares dropped sharply. If the price of the shares had increased, the investor would have lost money.

Table 9.3 Cash Flows from the Purchase of Shares and the Shorting of Shares

Time	Cash Flows if Shares are Purchased (USD)	Cash Flows if Shares are Shorted (USD)
Today	−5,000	+5,000
1 Month	+200	−200
2 Months	+4,000	−4,000
Total	−800	+800

Note that (assuming no borrowing fee) an investor who shorts the shares and closes out the position at the end of two months is in the opposite position of an investor who buys the shares and sells them at the end of the two months. This is illustrated in Table 9.3.

An investor who shorts is required to maintain a margin account (consisting of cash and/or marketable securities) with his or her broker. The margin account ensures the investor will not walk away from the agreement if the share price increases. The initial margin is typically 50% of the value of the shares plus the proceeds from selling the shares. This may be increased if the share price increases.

Many other assets, such as bonds, can be shorted like stocks. The investor can maintain the short position for as long as the asset can be borrowed. Occasionally, an investor is forced to close out a short position because the asset can no longer be borrowed.

9.4 THE LAW OF ONE PRICE AND ARBITRAGE

The law of one price states that if two portfolios provide the same future cash flows, they should sell for the same price. If the law of one price did not hold, there would be theoretical arbitrage opportunities.[2]

To see why this is the case, suppose that:

- The price of Portfolio X is greater than the price of Portfolio Y, and
- They will provide the same cash flows at the same time in the future.

[2] Arbitrage is the simultaneous purchase and sale of assets to profit from price differences. It is a trade that profits by exploiting the price differences of identical or similar financial instruments in different markets or in different forms.

Here is an arbitrage opportunity. Consider first a trader who owns the more expensive portfolio (Portfolio X). He or she could sell Portfolio X and buy Portfolio Y for an immediate cash inflow while still receiving the same future cash flows. Consider next a trader who does not own either portfolio. This trader can simply short Portfolio X and buy Portfolio Y; the trader will lock in a profit as long as the fee for borrowing Portfolio Y is not too great.

The existence of traders pursuing arbitrage opportunities will usually cause market prices to move until the existence of the arbitrage opportunity is eliminated. In our example, the price of Portfolio X will decline and the price of Portfolio Y will increase as traders sell Portfolio X and buy Portfolio Y; this will continue until the arbitrage profit disappears. The law of one price then holds.

To illustrate the law of one price, we will consider the securities in Table 9.1 and 9.2 with a maturity date of May 31, 2018. Consider first the Treasury bill in Table 9.1. This has a quoted bid price of 1.630 and a quoted ask price of 1.620. The number of calendar days between March 9, 2018 and May 31, 2018 is 83. Using Equation (9.1), the bid price of the Treasury bill is

$$100 - \frac{1.630 \times 83}{360} = 99.6242$$

The ask price is

$$100 - \frac{1.620 \times 83}{360} = 99.6265$$

The discount factor calculated from the bid price is therefore 0.996242, and the discount factor calculated from the ask price is 0.996265.

Consider next the first of the three bonds in Table 9.2 with a May 31, 2018 maturity date. The last coupon payment date was November 30, 2017 and there are 182 days between November 30, 2017 and May 31, 2018. The settlement date is March 12, 2018 and there are 102 days between November 30, 2017 and March 12, 2018. Because the coupon is 0.875% per year, the accrued interest is

$$0.5 \times 0.875 \times \frac{102}{182} = 0.2452$$

The bid quoted price is USD 99.8125. The bid cash price for the bond is therefore USD 100.0577 (= 99.8125 + 0.2452). The ask quoted price is USD 99.8281. The ask cash price is therefore USD 100.0733 (= 99.8281 + 0.2452). The cash amount received at maturity is USD 100.4375 (= 100 + (0.5 × 0.875)). The discount factor calculated from the bid price is therefore:

$$\frac{100.0577}{100.4375} = 0.996218$$

The discount factor calculated from the ask price is

$$\frac{100.0733}{100.4375} = 0.996374$$

Similar calculations can be carried out for the other two bonds in Table 9.2 maturing on May 31, 2018. The results are summarized in Table 9.4. The bid and ask discount factors are all close to each other. In this case, all the ask discount factors are greater than all the bid discount factors. This indicates the four instruments we are considering do not provide any arbitrage opportunities for investors. Thus, it was not possible to lock in a profit by buying a Treasury instrument at the ask price and selling an equivalent instrument at the bid price. This provides support for the law of one price.

However, we should not assume there are never any arbitrage opportunities in Treasury markets. This is because factors other than promised cash flows are occasionally considered in the way Treasury instruments are priced. For example, the tax treatment of a Treasury instrument and its liquidity can be important in determining price. We now consider how violations of the law of one price are sometimes created by liquidity issues.

Liquidity

Liquidity is a measure of how actively a financial instrument trades. Consider two bonds promising the same future cash flows. Bond X trades actively and therefore is highly liquid, whereas Bond Y is relatively illiquid and perhaps trades only a few times a year. It is likely Bond X will have a higher price than Bond Y. This is because a trader buying Bond X knows it will not be difficult to sell the bond should the need arise. In contrast, a trader buying Bond Y is less certain about how easy it would be to sell the bond in the future.

Table 9.4 Alternative Treasury Investments Maturing on May 31, 2018

Instrument	Bid Cash Price	Ask Cash Price	Final Cash Flow	Bid Discount Factor	Ask Discount Factor
Treasury bill	99.6242	99.6265	100	0.996242	0.996265
0.875% Coupon Treasury bond	100.0577	100.0733	100.4375	0.996218	0.996374
1% Treasury bond	100.1161	100.1318	100.5000	0.996180	0.996337
2.375% Treasury bond	100.8061	100.8218	101.1875	0.996231	0.996386

An arbitrageur can buy Bond Y and sell (or short) Bond X. This will generate a positive cash flow initially due to the difference in price. Later on, the cash flows eventually received from the positions in the two bonds will cancel each other out. This is a simplified version of the strategy followed by hedge fund Long Term Capital Management (LTCM). Simply put, the portfolio bought is less liquid than the portfolio sold. This strategy is called a convergence arbitrage because the prices of two portfolios are expected to converge to the same value when they promise the same cash flows.

LTCM was very successful for a few years before going bankrupt in 1998. Its problems arose when Russia defaulted on its debt, creating a flight to quality in debt markets. Liquid instruments saw their values soar relative to non-liquid instruments. LTCM was highly levered and could not meet its margin calls.

LTCM had used the law of one price to set up arbitrage positions. If it had been able to hold those positions until maturity, it probably would have been fine. But the firm and its positions were too highly levered to survive unexpected short-term market movements. The LTCM story shows that being right in the long term is not enough for a successful arbitrage; the arbitrageur must also be able to survive unexpected short-term price changes.

9.5 DISCOUNT FACTORS FROM COUPON-BEARING BONDS

We now show how discount factors can be obtained from coupon-bearing bonds. The approach is similar to the bootstrap method introduced in Chapter 19 of Financial Markets and Products. However, this method applies to discount rates rather than interest rates.

We will focus on the bonds listed in Table 9.5 (which are a subset of the bonds in Table 9.2). To make the analysis simpler, we will work with mid-market prices (i.e., the average of the bid and ask prices). Table 9.5 shows the mid-market dirty price.

Consider the first bond in Table 9.5. The bid and ask clean prices from Table 9.2 are USD 100.8906 and USD 100.9063.

Table 9.5 Selected Coupon-Bearing Bonds

Bond	Cash (Dirty) Price
4% maturing on Aug. 15, 2018	101.1747
2.75% maturing on Feb. 15, 2019	100.8071
0.75% maturing on Aug. 15, 2019	98.0440
3.625% maturing on Feb. 15, 2020	102.7581

Table 9.6 Discount Factors on March 9, 2018 Calculated from the Bond Data in Table 9.5

Maturity	Discount Factor
August 15, 2018	0.991909
February 15, 2019	0.980944
August 15, 2019	0.969406
February 15, 2020	0.956909

The mid-market clean price is the average of these two, or USD 100.8985. The accrued interest is USD 0.2762 and therefore the mid-market dirty price is USD 101.1747. Similar calculations are carried out for the other three bonds.

Denote $d(1)$, $d(2)$, $d(3)$, and $d(4)$ as the discount factors for August 15, 2018; February 15, 2019; August 15, 2019; and February 15, 2020 (respectively).

The first bond can be used in a straightforward way to find $d(1)$. The bond provides a final cash flow of USD 102 ($= 100 + (0.5 \times 4)$) and no intermediate cash flows. The discount factor is therefore:

$$d(1) = \frac{101.1747}{102} = 0.991909$$

The second bond provides a cash flow of USD 1.375 ($= 0.5 \times 2.75$) on August 15, 2018 and a final payment of USD 101.375 ($= 100 + (0.5 \times 2.75)$) on February 15, 2019. The value of the bond is therefore:

$$d(1) \times 1.375 + d(2) \times 101.375$$

We have already calculated $d(1)$, and we know that the value of the bond is 100.8071. To find $d(2)$, we can therefore solve

$$0.991909 \times 1.375 + d(2) \times 101.375 = 100.8071$$

This gives

$$d(2) = \frac{100.8071 - 0.991909 \times 1.375}{101.375} = 0.980944$$

Similarly:

$$0.991909 \times 0.375 + 0.980944 \times 0.375 + d(3) \times 100.375 = 98.0440$$

This gives $d(3) = 0.969406$. In the same way it can be shown that $d(4) = 0.956909$. These discount factors are summarized in Table 9.6.

The discount factor is a declining function of maturity (due to the time-value of money phenomenon).[3]

[3] Note that the discount rate is not a declining function of maturity when interest rates are negative. Since the 2007–2008 financial crisis, we have seen negative interest rates on the euro, Swiss franc, and Japanese yen.

Replicating Bond Cash Flows

Another bond maturing on February 15, 2020 in Table 9.2 is the 8.5% coupon bond that is bid USD 111.8906 and ask USD 111.9063, with a mid-market clean price of USD 111.8985 (= 0.5 × (111.8906 + 111.9063)). By adding accrued interest, we see the dirty (cash) price is USD 112.4855.

An analyst might be interested in whether this bond is mispriced. One way of investigating this is by replicating the bond's cash flows using the bonds in Table 9.5. The bond's cash flows are as follows:

August 15, 2018:	USD 4.25
February 15, 2019:	USD 4.25
August 15, 2019:	USD 4.25
February 15, 2020:	USD 104.25

Of the bonds in Table 9.5, the only one that gives a cash flow at time February 15, 2020 is the fourth one. It provides a cash flow of 101.8125 at this date. If we take a position of:

$$\frac{104.25}{101.8125} = 1.023941$$

in the bond, we replicate the final cash flow. Now consider the August 15, 2019 date. The cash flow we must replicate on that date is USD 4.25. The position just calculated in the fourth bond gives us a cash flow of USD 1.8559 (= 1.023941 × 1.8125). We therefore need an additional cash flow of USD 2.3941 (= 4.25 − 1.8559). Of the three remaining bonds, the only one that gives a cash flow on August 15, 2019 is the third one. Its cash flow on that date is USD 100.375. We therefore need a position in that bond of:

$$\frac{2.3941}{100.375} = 0.023852$$

Next, we consider the February 15, 2019 date. The third and fourth bonds provide a cash flow of USD (1.8648 = (1.023941 × 1.8125) + (0.023852 × 0.375)). We therefore need an additional cash flow of USD 2.3852 (= 4.25 − 1.8648). The position needed in the second bond is therefore:

$$\frac{2.3852}{101.375} = 0.023528$$

Similarly, the position in the first bond is USD 0.023066.

Table 9.7 summarizes these calculations and shows that the cost of replicating the 8.5% coupon February 15, 2020 bond is USD 112.26232, which is slightly different from the USD 112.4855 mid-market price of the bond. This suggests that a potential arbitrage would involve shorting the bond and replicating it as in Table 9.7 (but it is quite likely that bid-ask spreads would eliminate the arbitrage opportunity in practice). The numbers we have calculated are for bonds with a face value of USD 100. When carrying out the arbitrage, we might replicate a USD 10 million position in the bond and short USD 10 million face value of the bond.

Note that determining discount factors from coupon-bearing bonds involves starting with the shortest maturity bond and looking at progressively longer maturity bonds. Replicating the cash flows of a bond in the way we have just described involves the reverse; we start by replicating the longest maturity cash flow and work backwards.

9.6 STRIPS

STRIPS is an acronym for Separate Trading of Registered Interest and Principal of Securities. STRIPS are created by investment dealers when a coupon-bearing bond is delivered to the Treasury and exchanged for its principal and coupon components. The securities created from the coupon payments are known as TINTs, INTs, or C-STRIPS. The securities created from principal payments are known as TPs, Ps, or P-STRIPS.

Consider the May 15, 2030 bond paying a coupon of 6.25% in Table 9.2. A bond with a face value of USD 1,000,000 provides a coupon of USD 31,250 every May 15 and November 15 and a final principal payment of USD 1,000,000. This would be converted into 25 C-STRIPS and one P-STRIP as indicated in Table 9.8.

Table 9.7 Replicating Portfolio for the 8.5% Coupon 2020 Bond

Bond	Position	Cost of Bond	Cost of Position
4% maturing on Aug. 15, 2018	0.023066	101.1747	2.33373
2.75% maturing on Feb. 15, 2019	0.023528	100.8071	2.37182
0.75% maturing on Aug. 15, 2019	0.023852	98.0440	2.33853
3.625% maturing on Feb. 15, 2020	1.023941	102.7581	105.21824
Total			112.26232

Table 9.8 Creation of C-STRIPS and P-STRIPS from the May 15, 2030 6.25% Bond on March 9, 2018

Date	C-STRIP Face Value (USD)	P-STRIP Face Value (USD)
May 15, 2018	31,250	
November 15, 2018	31,250	
May 15, 2019	31,250	
November 15, 2019	31,250	
....		
November 15, 2029	31,250	
May 15, 2030	31,250	1,000,000

Table 9.9 gives a sample of some C-STRIP and P-STRIP prices on March 9, 2018. The prices provide direct estimates of discount rates. For example, the November 15, 2047 P-STRIP is bid USD 38.945 and ask USD 39.059. The mid-market discount factor for that date is therefore 0.39002 (= 0.5 × (0.38945 + 0.39059)).

The table shows that (in theory) there were some arbitrage opportunities available on March 9, 2018. For example, one could buy the May 15, 2021 C-STRIP for USD 92.426 and short the May 15, 2021 P-STRIP for USD 92.499. However, transactions costs would probably eliminate any gains from this position.

9.7 DAY-COUNT CONVENTIONS

Day-count conventions describe the way in which interest is earned through time. We usually know the amount of interest earned over a reference period (e.g., the time between coupon payments or over a whole year) and we want to know the interest earned over a particular holding period.

A day-count convention is usually expressed as *X*/*Y*. *X* defines how the holding period time interval is measured; *Y* defines how the length of the reference period is measured. The interest earned during the holding period is calculated as

$$\frac{\text{Number of days in holding period}}{\text{Number of days in reference period}} \times \begin{array}{c}\text{Interest earned in}\\ \text{reference period}\end{array}$$

Three common day-count conventions are

- Actual/actual (in period),
- 30/360, and
- Actual/360.

The actual/actual (in period) day-count convention applies to U.S. Treasury bonds and is the convention used in Equation (9.2).

Table 9.9 P-STRIP and C-STRIP Prices on March 9, 2018

Maturity	Bid	Ask
P-STRIPS		
May 15, 2020	95.367	95.387
May 15, 2021	92.499	92.528
Nov. 15, 2022	88.634	88.675
Aug. 15, 2023	86.497	86.544
Nov. 15, 2024	83.143	83.197
Aug. 15, 2025	81.216	81.275
Nov. 15, 2026	78.089	78.156
Nov. 15, 2027	75.612	75.684
Nov. 15, 2028	73.271	73.348
Aug. 15, 2029	71.675	71.756
May 15, 2030	70.100	70.184
Feb. 15, 2031	68.627	68.714
Feb. 15, 2036	58.712	58.816
May 15, 2037	56.272	56.378
May 15, 2038	54.361	54.469
Nov. 15, 2039	51.373	51.483
Nov. 15, 2040	49.457	49.568
Nov. 15, 2041	47.659	47.770
Nov. 15, 2042	45.965	46.076
Nov. 15, 2043	44.345	44.458
Nov. 15, 2044	42.902	43.014
Nov. 15, 2045	41.507	41.621
Nov. 15, 2046	40.143	40.256
Nov. 15, 2047	38.945	39.059
C-STRIPS		
May 15, 2020	95.238	95.259
May 15, 2021	92.397	92.426
May 31, 2021	92.365	92.395
Nov. 15, 2024	82.719	82.774
Nov. 15, 2027	74.823	74.895
May 15, 2030	68.974	69.057
Aug. 15, 2030	68.253	68.337
Nov. 15, 2030	67.705	67.790
Feb. 15, 2036	57.102	57.203
Nov. 15, 2040	48.745	48.854
Nov. 15, 2047	38.831	38.945

Consider a holding period that is only part of the period between two coupon payment dates. The fraction of the next coupon payment earned by the investor is the holding period divided by the number of actual days in the period between coupon payment dates.

The 30/360 day-count convention is used for corporate and municipal bonds in the United States. This means calculations are carried out assuming 30 days in a month and 360 days in a year. For example, suppose coupons with a rate of 10% per year are paid on March 5 and September 5, and that we want to know the amount of interest earned between March 5 and June 10. The convention assumes 30 days in March, April, and May (so that there are 95 days between March 5 and June 10). The interest earned on USD 1,000 in a year is 10% of USD 1,000 (or USD 100). The USD interest earned between March 5 and June 10 is

$$\frac{95}{360} \times 100 = 26.39$$

An interesting consequence of the 30/360 day-count convention is that three days of interest are earned between February 28 and March 1 (in both leap years and non-leap years). Meanwhile, no interest is earned on the 31^{st} of a month.

The actual/360 day-count convention is used for money market instruments in the United States (as we saw this in Section 9.1). The quoted interest is per 360 days and the actual number of days in a holding period is used to calculate the interest earned during the holding period. As we saw in Section 9.1, a further complication is that the interest is calculated as a percentage of the face value (not as a percentage of the price of the bond).

Conventions vary from country to country. For example, money market instruments are quoted on an actual/365 basis in Canada, Australia, and New Zealand.

SUMMARY

Discount factors describe the relationship between the value of a cash flow received at some time in the future and the value of the cash flow today. They can be calculated in a direct way from the prices of Treasury bills, P-STRIPS, and C-STRIPS.

For a set of coupon-bearing bonds, it is necessary to list the bonds in order of maturity. We start with the shortest maturity bond and work down the list by calculating discount factors for progressively longer maturities. When the nth bond in the list is considered, the discount factors for all payments except the final one have already been calculated. Thus, the discount factor for the final payment can be determined from the price of the bond.

The law of one price states that two portfolios providing the same cash flows at the same time in the future should be the same price. If they are not the same price, there is an arbitrage opportunity. A trader can short the higher priced portfolio and buy the lower priced portfolio. In practice, there are a number of factors that can lead to pricing discrepancies. These include relative liquidity, taxation concerns, and transaction costs.

The cash flows of a bond can be replicated from the cash flows of other bonds maturing on the bond's payment dates. This enables an analyst to check the consistency of the pricing of different bonds.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

QUESTIONS

Short Concept Questions

9.1 How are Treasury bill prices quoted?

9.2 What is the maximum life of a Treasury bill?

9.3 How is the accrued interest on a Treasury bond calculated?

9.4 What is the difference between the cash price and quoted price of a Treasury bond?

9.5 Explain the difference between the clean price and dirty price of a Treasury bond.

9.6 What is the law of one price?

9.7 Give two reasons why the law of one price might not hold for Treasury bonds.

9.8 Explain how discount factors are calculated from coupon-bearing bond prices.

9.9 What are (a) C-STRIPS and (b) P-STRIPS?

9.10 How many days interest are earned between May 30 and June 1 using the 30/360 day-count convention?

Practice Questions

9.11 A U.S. Treasury bill lasts 35 days and has a quoted price of 1.40. What is the cash price? What is the 35-day discount factor?

9.12 The price of a U.S. Treasury bond is quoted as 98.0. It is sold in a transaction settled on June 27. Coupons are paid at the rate of 6% per year on March 15 and September 15. What is the cash price?

9.13 A six-month bond that pays coupons at the rate of 5% per year is currently worth 100.5. What is the six-month discount factor?

9.14 In addition to the bond in Question 9.13, a one-year bond that pays coupons every six months at the rate of 3% per year is currently worth 98.5. What is the one-year discount factor?

9.15 Can discount factors ever be an increasing function of maturity? Discuss.

9.16 You can trade bonds lasting 0.5 years and one year that have coupons of 3% and 4%, respectively. If these bonds pay their coupons on a semi-annual basis, how could you use them to replicate the cash flows on a one-year bond paying a 5% coupon?

9.17 If the discount factors for six months, 12 months, 18 months, and 24 months are 0.99, 0.98, 0.97, and 0.96, (respectively). What is the price of a two-year bond paying a coupon of 3% per year (on a semi-annual basis)?

9.18 Suppose a U.S. Treasury bond that pays coupons at the rate of 8% per year on May 15 and November 15 is sold in a transaction settled on October 18. What is the accrued interest?

9.19 What difference would it make to your answer for Question 9.18 if the bond were a corporate bond with a 30/360 day-count convention?

9.20 Why do traders prefer to quote clean prices rather than dirty prices for Treasury bonds?

The following questions are intended to help candidates understand the material. They are <u>not</u> actual FRM exam questions.

ANSWERS

Short Concept Questions

9.1 Treasury bill prices consider the interest over a 360-day period. The interest is as a percent of the face value, not the amount invested.

9.2 One year.

9.3 Accrued interest is calculated by dividing the number of calendar days from the previous coupon payment date by the number of calendar days between coupon payment dates and multiplying the result by the amount of the previous coupon.

9.4 The cash price is the quoted price plus accrued interest.

9.5 The clean price is the quoted price. The dirty price is the cash price and it equals the clean price plus the accrued interest.

9.6 The law of one price states that two portfolios that will provide the same cash flows in the future (with the cash flows having the same timing) should sell for the same price.

9.7 Liquidity and taxes. As a bond becomes more liquid, its price tends to increase. A bond will also have a greater price than another bond if it has a more favorable tax treatment.

9.8 We work from the shortest maturity bond to the longest maturity bond calculating discount factors. A bond maturing at time T can be used to calculate the discount factor for that maturity because the discount factors relevant to valuing its coupons have already been calculated.

9.9 C-STRIPS and P-STRIPS are created by stripping the coupon and principal payments from a bond and selling them separately. C-STRIPS are created from the coupons; P-STRIPS are created from the principal payments.

9.10 There are two calendar days, but only one day's interest is earned.

Solved Problems

9.11 The cash price is

$$100 - \frac{1.40 \times 35}{360} = 99.8639$$

The discount factor for 35 days is 0.998639.

9.12 There are 184 days between coupon payments and 104 days between the last coupon and the settlement date. The accrued interest is

$$3 \times \frac{104}{184} = 1.6957$$

The cash price is $98 + 1.6957 = 99.6957$.

9.13 The discount factor is $100.5/(100 + 2.5) = 0.980488$.

9.14 The value of the coupon that will be paid in six months is $1.5 \times 0.980488 = 1.4707$. This means that the value of 101.5 received in one year is $98.5 - 1.4707 = 97.0293$ so that the one-year discount factor is $97.0293/101.5 = 0.955953$.

9.15 Discount factors normally decrease as the time to maturity increases. An exception would be when interest rates are negative. Since the 2007–2008 crisis interest rates have been negative in several countries.

9.16 A one-year 5% coupon bond provides cash flows of 2.5 in six months and 102.5 in one year. We first replicate the one-year cash flow with:

$$\frac{102.5}{102} = 1.0049$$

of the one-year bond. This will provide a cash flow of $1.0049 \times 2 = 2.0098$ at the six-month point. We therefore require an extra 0.4902 of cash flow at this point. This can be provided by:

$$\frac{0.4902}{101.5} = 0.00483$$

of the six-month bond.

9.17 The bond price is

$$0.99 \times 1.5 + 0.98 \times 1.5 + 0.97 \times 1.5 + 0.96 \times 101.5 = 101.85$$

9.18 There are 184 days between coupon payments and 156 days between the last coupon and the settlement date. The accrued interest is

$$4 \times \frac{156}{184} = 3.3913$$

9.19 In this case, there are 153 days between the last coupon and the settlement date. The accrued interest is

$$8 \times \frac{153}{360} = 3.4000$$

9.20 Clean prices are less variable than dirty prices. They reflect term structure changes but do not reflect accrued interest, which shows a pronounced saw-tooth pattern.

10

Interest Rates

Learning Objectives

After completing this reading, you should be able to:

- Calculate and interpret the impact of different compounding frequencies on a bond's value.
- Define spot rate and calculate discount factors given spot rates.
- Interpret the forward rate and calculate forward rates given spot rates.
- Define par rate and describe how to determine the par rate of a bond.
- Interpret the relationship between spot, forward, and par rates.
- Assess the impact of a change in time to maturity on the price of a bond.
- Define the "flattening" and "steepening" of rate curves and describe a trade to reflect expectations that a curve will flatten or steepen.
- Describe a swap transaction and explain how a swap market defines par rates.

In the previous chapter, we showed how the time value of money can be described using discount factors. If we know the discount factors for all future times, we can value any Treasury instrument. Conversely, we can imply discount factors if we know the market prices of a range of different Treasury instruments.

Investors and traders prefer to express the time value of money in terms of interest rates rather than discount factors. To understand an interest rate, we need to understand the compounding frequency used to measure it. We also need to understand the different types of interest rates. This chapter discusses spot rates, forward rates, and par rates. It also explains overnight rates and swap rates.[1]

10.1 MEASURING INTEREST RATES

To fully describe an interest rate, we need to specify the compounding frequency with which it is measured. An interest rate of 8% per annum with annual compounding would mean that USD 100 will grow to 108 (= 100 × 1.08) at the end of one year. It will grow to:

$$100 \times 1.08^2 = 116.64$$

at the end of two years because the USD 108 is invested at 8% for a second year. More generally, it will grow to:

$$100 \times 1.08^n$$

at the end of n years.

Now suppose the 8% per annum is measured with semi-annual compounding. This means that 4% is earned every six months. USD 100 will grow to:

$$100 \times 1.04^2 = 108.16$$

at the end of one year. It will grow to:

$$100 \times 1.04^4 = 116.99$$

at the end of two years and:

$$100 \times 1.04^{2n}$$

at the end of n years.

If you are lending money, you would clearly prefer the specified interest rate to be compounded semi-annually rather than annually. If you are borrowing money, however, you would prefer it to be annually compounding.

[1] There is some overlap between the material in this chapter and the material in the Interest Rates chapter of Financial Markets and Products. However, the concepts presented there are so important to fixed-income markets that for completeness they are covered again here in the context of those markets.

We can continue this example. If the interest rate is compounded quarterly, an investor earns 2% every quarter (with the proceeds being compounded in the next quarter). The amount to which USD 100 grows at the end of n years is then:

$$100 \times 1.02^{4n}$$

Table 10.1 shows how USD 100 at an 8% rate grows at various compounding frequencies. In understanding Table 10.1, the key point is that the compounding frequency defines the units of measurement for an interest rate. The difference between annual compounding and semi-annual compounding is analogous to the difference between measuring temperature in degrees Fahrenheit or degrees centigrade. For example, an interest rate of 8.16% measured with annual compounding is equivalent to 8% with semi-annual compounding.

Knowing the compounding frequency is also important when present values are calculated. Table 10.2 shows the present value of USD 100 received in five years when the interest rate is 8% for several different compounding frequencies. Consider, for example, the quarterly row. The rate used for discounting is 2% per three months. There are 20 three-month periods in five years, and so the present value of USD 100 in five years is

$$\frac{100}{(1 + 0.02)^{20}} = 67.30$$

The compounding frequency used for an interest rate is often the same as the frequency of payments, but this is not always the case. For example, the interest rate on a Canadian fixed-interest mortgage is expressed with semi-annual compounding even though payments are made every month or every two weeks. In the United States, this mortgage interest rate would be expressed with monthly compounding, while in the United Kingdom it would be expressed with annual compounding. These ways of expressing interest rates are government requirements designed to make it easier for borrowers to compare interest rates from different lenders.

Suppose rate R_1 is compounded m_1 times per annum, and we want to calculate the equivalent rate compounded m_2 times per annum (written as R_2). The R_1 rate means that an amount A compounds to:

$$A\left(1 + \frac{R_1}{m_1}\right)^{m_1} \qquad (10.1)$$

at the end of one year. The R_2 rate means that an amount A compounds to:

$$A\left(1 + \frac{R_2}{m_2}\right)^{m_2}$$

Table 10.1 Value to Which USD 100 Grows in One Year as the Compounding Frequency for the 8% Rate is Increased

Compounding Frequency	Number of Times Rate Is Compounded per Year	Value to Which USD 100 Grows in One Year
Annual	1	108.00
Semi-annual	2	108.16
Quarterly	4	108.24
Monthly	12	108.30
Weekly	52	108.32
Daily	365	108.33

Table 10.2 Present Value of USD 100 Received in Five Years as the Compounding Frequency of the 8% Rate Used for Discounting is Increased

Compounding Frequency	Number of Times Rate is Componded per Year	Present Value of USD 100 Received in 5 Years
Annual	1	68.06
Semi-annual	2	67.56
Quarterly	4	67.30
Monthly	12	67.12
Weekly	52	67.05
Daily	365	67.03

at the end of one year. Rate R_2 compounded m_2 times per year is therefore equivalent to rate R_1 compounded m_1 times per year when:

$$A\left(1+\frac{R_1}{m_1}\right)^{m_1} = A\left(1+\frac{R_2}{m_2}\right)^{m_2}$$

or

$$R_2 = \left[\left(1+\frac{R_1}{m_1}\right)^{m_1/m_2} - 1\right]m_2$$

To illustrate this formula, suppose the rate is 5% with semi-annual compounding, and we wish to calculate the equivalent rate with quarterly compounding. In this case $m_1 = 2$, $m_2 = 4$, and $R_1 = 0.05$:

$$R_2 = \left[\left(1+\frac{0.05}{2}\right)^{2/4} - 1\right] \times 4 = 0.04969$$

The equivalent rate with quarterly compounding is therefore 4.969%.

Continuous Compounding

In Tables 10.1 and 10.2, we consider daily compounding (i.e., compounding 365 times per year). What happens as we increase the compounding frequency still further? For example, we could compound every hour or every minute or even every second. In the limit, we obtain what is referred to as continuous compounding. If an interest rate R is continuously compounded, it can be shown that an amount A grows to:

$$Ae^{RT} \tag{10.2}$$

by time T, where e is the mathematical constant approximately equal to 2.71828. The present value of an amount A received at time T is

$$Ae^{-RT}$$

If we extend Tables 10.1 to consider continuous compounding, USD 100 invested at a rate of 8% measured grows to:

$$100e^{0.08\times 1} = 108.33$$

Similarly, if we extend Table 10.2 to consider continuous compounding, the present value of USD 100 received in five years when the discount rate is 8% is

$$100e^{-0.08\times 5} = 67.03$$

In both cases, the continuously compounded rate is the same as the daily compounded rate when rounded to two decimal places.

When we move on to consider derivatives such as options, we will use interest rates measured with continuous compounding. It therefore makes sense to get used to working with continuously compounded interest rates now. In fact, by comparing Equations (10.1) and (10.2), we see that they lead to simpler formulas than rates compounded m times per year.

To convert a rate expressed with compounding frequency m times per year to an equivalent continuously compounded rate, we equate the values at the end of a year. Define R_m as a rate expressed with compounding m times per year and R_c as the equivalent continuously compounded rate. We must have

$$Ae^{R_c\times 1} = A\left(1 + \frac{R_m}{m}\right)^m$$

so that:[2]

$$R_c = m\ ln\left(1 + \frac{R_m}{m}\right)$$

and

$$R_m = m(e^{R_c/m} - 1)$$

Suppose a rate is 5% when expressed with semi-annual compounding. The equivalent rate with continuous compounding is

$$2\ ln\left(1 + \frac{0.05}{2}\right) = 0.0494$$

or 4.94%. As a further example, suppose the rate is 6% when expressed with continuous compounding. The equivalent rate with quarterly compounding is

$$4(e^{0.06/4} - 1) = 0.0605$$

or 6.05%.

10.2 SPOT RATES

The spot rate is the interest rate earned when cash is received at just one future time. It is also referred to as the zero-coupon interest rate, or just the "zero."

[2] The natural logarithm function ln is the inverse of the exponential function.

Suppose you invest USD 100 today and are repaid with USD 120 in three years with no intermediate payments. The three-year spot rate is the rate that equates USD 120 in three years with USD 100 today. If the rate R is measured with annual compounding, it is given by solving

$$100(1 + R)^3 = 120$$

In this case, R is 6.27%. We could convert this to some other compounding frequency using the formulas derived in the previous section. Alternatively, we could build the compounding frequency into the original calculation. With semi-annual compounding for example, the rate is given by:

$$100\left(1 + \frac{R}{2}\right)^6 = 120$$

so that it becomes 6.17%.

Spot rates give the same information as discount factors (which were introduced in Chapter 9). Suppose the discount factor for t years is $d(t)$ and that the t-year spot rate is $r(t)$ with semi-annual compounding (this is the usual compounding frequency for Treasury bonds). An investment of 100 will grow to:

$$100\left(1 + \frac{r(t)}{2}\right)^{2t}$$

When the discount factor $d(t)$ is applied to this, it should bring it back to 100. Hence:

$$100\left(1 + \frac{r(t)}{2}\right)^{2t} d(t) = 100$$

so that:

$$d(t) = \left(1 + \frac{r(t)}{2}\right)^{-2t}$$

If the spot rate had been expressed with continuous compounding, we would have

$$d(t) = e^{-r(t)t}$$

10.3 PAR RATES

In the previous chapter, we noted Treasury bonds pay coupons every six months at a certain annual rate (e.g., when the coupon rate is 7%, 3.5% of the face value is paid every six months).

Consider a bond with maturity T. When the coupon rate is zero, we have the equivalent of a stripped bond and thus the value of the bond is less than its face value.[3] As the coupon rate is increased, the value of the bond increases. For some value of the coupon rate, the value of the bond will be equal to its face value. This is referred to as the par rate.

[3] This assumes the interest rate is not negative.

The payment dates on the bonds we are considering are 0.5, 1, 1.5, . . . , T years. Suppose the par rate is p. From the definition of par rate that we have just given, a bond paying $p/2$ at times 0.5, 1, 1.5, . . . , T years, and 100 at time T is worth 100. This means that:

$$\frac{p}{2}d(0.5) + \frac{p}{2}d(1) + \frac{p}{2}d(1.5) + \ldots + \frac{p}{2}d(T) + 100d(T) = 100 \quad (10.3)$$

where $d(t)$ is the discount factor for a maturity of t years.[4]

Define $A(T)$ as the value of an instrument that pays USD 1 on every payment date (this is referred to as an annuity):

$$A(T) = d(0.5) + d(1.0) + d(1.5) + \cdots + d(T)$$

From Equation (10.3):

$$\frac{p}{2}A(T) + 100d(T) = 100$$

so that:

$$p = \frac{2 \times 100 \times (1 - d(T))}{A(T)} \quad (10.4)$$

Suppose the discount factors for 0.5, 1.0, 1.5, and 2.0 years are 0.98038, 0.95181, 0.92184, and 0.88849 (respectively). These correspond to spot interest rates, expressed with semi-annual compounding, of 4%, 5%, 5.5%, and 6% (respectively). When $T = 2$:

$$A(T) = 0.98038 + 0.95181 + 0.92184 + 0.88849 = 3.74252$$

$$d(T) = 0.88849$$

The two-year par rate (%) is from Equation (10.4):

$$p = \frac{2 \times 100 \times (1 - 0.88849)}{3.74252} = 5.9592$$

Thus, a two-year bond paying a coupon semi-annually at a rate of 5.9592% per year is worth par.

We can use the par rate in conjunction with the annuity factors $A(T)$ to provide a way of valuing bonds with other coupons. Consider a bond with maturity T, coupon c, and a face value of 100. The value of the bond (V) is

$$V = \frac{c}{2}A(T) + 100d(T)$$

We know that:

$$\frac{p}{2}A(T) + 100d(T) = 100$$

[4] In this chapter and subsequent chapters, we assume that payments on bonds occur every 0.5 years. In practice, this is not quite true because the number of days between coupon payments is not quite 0.5 years. Also, the conventions concerning holidays have to be considered.

Substituting for $d(T)$ we get

$$V = 100 + \frac{c - p}{2}A(T)$$

In our earlier example, the par rate is 5.9592%. When the coupon is 4%, this formula gives the value of the bond as:

$$100 + \frac{4 - 5.9592}{2} \times 3.74252 = 96.33$$

10.4 FORWARD RATES

Forward rates are the future spot rates implied by today's spot rates. For example, suppose the offered one-year rate is 3% and the offered two-year rate is 4% (both with annual compounding). As a rough approximation, we can say that the rate being offered for the second year is 5%. This is because 3% for the first year, when averaged with 5% for the second year, gives 4% for the two years.

Let us look at this a little more formally. Suppose F is the forward rate for the second year. The forward rate is such that USD 100, if invested at 3% for the first year and at a rate of F for the second year, gives the same outcome as 4% for two years. This means that:

$$100 \times 1.03 \times (1 + F) = 100 \times 1.04^2$$

so that:

$$F = \frac{1.04^2}{1.03} - 1 = 0.0501$$

or 5.01%. This is close to the 5% given by our approximate argument. However, it is not the same because there are non-linearities when rates are expressed with annual compounding.

When rates are expressed with semi-annual compounding (as is frequently the case in fixed-income markets), an extension of this analysis shows that the forward rate per six months for a six-month period starting at time T is

$$\frac{(1 + R_2/2)^{T+0.5}}{(1 + R_1/2)^T} - 1 \quad (10.5)$$

where R_1 and R_2 are the spot rates for maturities T and $T + 0.5$ (respectively) with semi-annual compounding. Thus, the annualized forward rate expressed with semi-annual compounding is twice this.

Interestingly, when rates are expressed with continuous compounding, the non-linearities disappear and the averaging argument gives an exact answer. To slightly change our example, suppose that the offered one-year rate is 3% and the offered two-year rate is 4% (both with continuous compounding). The forward rate F, when expressed with continuous compounding, must satisfy

$$100e^{0.03\times 1}e^{F\times 1} = 100e^{0.04\times 2}$$

In this case, F is exactly equal to 0.05 (or 5%).

When rates are expressed with continuous compounding, the forward rate for the period between time T_1 and T_2 is

$$F = \frac{R_2T_2 - R_1T_1}{T_2 - T_1}$$

where R_1 is the spot rate for maturity T_1 and R_2 is the spot rate for maturity T_2. This formula is approximately true when other compounding frequencies are used for the rates.

Suppose that the spot rate for 5 years is 5% and the spot rate for 5.5 years is 5.2% (both with continuous compounding). The forward rate for the period between year 5 and year 5.5 is

$$\frac{0.052 \times 5.5 - 0.05 \times 5.0}{5.5 - 5.0} = 0.072$$

or 7.2% with continuous compounding. While this appears to be a high rate, note that 7.2% earned over half a year combined with 5% earned over 5 years averages to 5.2% (= ((7.2 × 0.5) + (5 × 5))/5.5) over 5.5 years.

When forward rates for successive periods are compounded, we get spot rates. For example, suppose all rates are expressed with semi-annual compounding, and that $F_1, F_2, \ldots F_n$ are forward rates in n successive six-month periods. Then:

$$\left(1 + \frac{R}{2}\right)^n = \left(1 + \frac{F_1}{2}\right)\left(1 + \frac{F_2}{2}\right)\ldots\left(1 + \frac{F_n}{2}\right) \quad (10.6)$$

If a large financial institution can borrow or lend at spot rates, it can lock in the forward rate. Suppose that in the example considered earlier, a financial institution wants to lock in the forward rate of 7.2% on its invested funds for the time between year five and year 5.5. It should borrow at 5% for five years and invest the funds at 5.2% for 5.5 years (all rates are quoted with continuous compounding). Suppose further that the amount borrowed is USD 100. This comes in and goes out at time zero for no net cash flow. At time five years, USD 128.40 (= $100e^{0.05\times 5}$) is used to repay the loan. At time 5.5 years, USD 133.11 (= $100e^{0.052\times 5.5}$) is received. The interest effectively earned over the six months is

$$133.11 - 128.40 = 4.71$$

on an investment of USD 128.40. The annualized rate with semi-annual compounding is 7.33%, and with continuous compounding it is 7.2% (as expected).

An instrument that guarantees that a particular rate will be earned (or paid) during a certain future period is called a forward rate agreement. When the guaranteed rate is the prevailing forward rate, the agreement is worth zero. This is because (as we have just explained) it is possible to enter into transactions that guarantee the forward rate for a future period. Suppose that a forward rate agreement guarantees that a rate R will be earned for a certain future period when the forward rate is F. The value of the forward rate agreement is the present value of $(R - F)$ applied to the principal amount. It is positive when $R > F$ and negative when $R < F$. We will see an application of this principle in the next section.

Maturity Effects

What happens to the price of a bond if the term structure remains unchanged over a six-month period? Suppose a bond lasts T years and provides a coupon of c. The cash flows are as indicated in Table 10.3. The difference between the original cash flows and the cash flows after six months are shown in the final row of the table. They are the cash flows from a forward rate agreement where interest is earned at rate c on 100 for six months.

As we saw in the previous section, the forward rate agreement has a positive value if the coupon is greater than the forward rate for the final period. In this case, the value of the cash flows after six months is less than their initial value because they are reduced by the value of this forward rate agreement. Similarly, the forward rate agreement has a negative value if the coupon is less than the forward rate for the final period. In this case, the value of the cash flows after six months is greater than their initial value.

Table 10.3 Cash Flows Where Term Structure Remains Unchanged

Time (Years)	0.5	1.0	1.5	. . .	T-1.0	T-0.5	T
Cash Flows Starting Today	c/2	c/2	c/2	. . .	c/2	c/2	100 + c/2
Cash Flows Starting in 6 Months	c/2	c/2	c/2	. . .	c/2	100 + c/2	
Difference						−100	100 + c/2

This shows that the value of a bond will rise (or fall) depending on whether the forward rate for the last period is greater (or less) than the coupon. In the case of an upward-sloping term structure, there will be a tendency for the forward rate to be higher than the coupon so that the bond price rises.

Trading Strategies

Suppose that it is expected that the actual rate realized for a future period between T_1 and T_2 will be less than the forward rate. (To keep the example simple, we suppose that an investor can borrow or lend at the same rate. This is approximately true for large financial institutions.) An investor can borrow funds for time T_1 and invest for time T_2. If the rate paid by the investor for the period between T_1 and T_2 is less than the forward rate, the investor's total financing cost will be less than the investor's return, producing a profit.

For example, suppose the two-year rate is 4% and the three-year rate is 5% (with both rates continuously compounded). The forward rate for the third year is 7%. If an investor feels confident that the rate for the third year will be less than 7%, the investor can borrow for two years at 4% and invest for three years at 5%. If the rate paid by the investor for the third year is in fact less than 7%, the investor's overall borrowing rate will be less than 5% and a profit will result.

Similarly, if the realized rate is expected to be greater than the forward rate, the investor should borrow for three years and invest for two years. If the investor is right, the rate that he or she is able to obtain for the third year will be greater than the forward rate of 7% and again a profit will result.

These examples show that forward rates play a key role in defining trading strategies.

10.5 PROPERTIES OF SPOT, FORWARD, AND PAR RATES

Key properties of the rates we have looked at so far are as follows.

- If the term structure is flat (with all spot rates the same), all par rates and all forward rates equal the spot rate.
- If the term structure is upward-sloping, the par rate for a certain maturity is below the spot rate for that maturity.
- If the term structure is downward-sloping, the par rate for a certain maturity is above the spot rate for that maturity
- If the term structure is upward-sloping, forward rates for a period starting at time T are greater than the spot rate for maturity T.
- If the term structure is downward-sloping, forward rates for a period starting at time T are less than the spot rate for maturity T.

The situation for an upward-sloping term structure case is illustrated in Table 10.4, while that for a downward-sloping term structure is illustrated in Table 10.5.

Table 10.4 Shown are Spot, Forward, and Par Rates for an Upward-Sloping Term Structure. All Rates are Semi-Annually Compounded. The Six-Month Forward is the Forward Rate for a Six-Month Period Starting on the Maturity Date

Maturity (yrs)	Spot	6-Mnth Fwd	Par
0.5	2.01	4.04	2.01
1	3.02	4.86	3.01
1.5	3.63	5.27	3.62
2	4.04	5.83	4.01
2.5	4.40	5.94	4.36
3	4.65	6.24	4.60
3.5	4.88	6.36	4.82
4	5.06	6.54	4.99
4.5	5.23	6.67	5.14
5	5.37		5.28

Table 10.5 Shown are Spot, Forward, and Par Rates for a Downward-Sloping Term Structure. All Rates are Semi-Annually Compounded. The Six-Month Forward is the Forward Rate for a Six-Month Period Starting on the Maturity Date

Maturity (yrs)	Spot	6-Mnth Fwd	Par
0.5	5.06	4.24	5.06
1	4.65	3.73	4.66
1.5	4.35	3.73	4.36
2	4.19	3.17	4.21
2.5	3.99	3.07	4.01
3	3.84	3.12	3.86
3.5	3.73	3.08	3.76
4	3.65	3.10	3.68
4.5	3.59	3.08	3.62
5	3.54		3.57

10.6 FLATTENING AND STEEPENING TERM STRUCTURES

A flattening term structure occurs

- When long- and short-maturity rates both move down, but long-maturity rates move down by more than short-maturity rates (known as a *bull flattener*); or
- When long- and short-maturity rates both move up, but short-maturity rates move up by more than long-maturity rates (known as a *bear flattener*).

A steepening term structure occurs

- When long- and short-maturity rates both move down but short-maturity rates move down by more than long-maturity rates; or
- When long- and short-maturity rates both move up, but short-maturity rates move up by less than long-maturity rates.

Note that a flattening term structure is not necessarily one that becomes flatter and a steepening term structure is not necessarily one that becomes steeper. If a term structure is already upward-sloping, then a steepening will cause it to be more upward-sloping and a flattening will cause it to be less upward-sloping. But if it is downward-sloping, the reverse is true. The steepening/flattening language refers to the relative rate movements and does not depend on the initial slope of the yield curve.

Suppose a trader thinks the current upward-sloping term structure will steepen so that 20-year rates increase faster than ten-year rates. The trader can short 20-year bonds and buy ten-year bonds. If the trader is right, the 20-year bonds will decline in value relative to the ten-year bonds and the trader will make money. Similarly, a trader who thinks that the term structure will flatten should buy 20-year bonds and short ten-year bonds.

10.7 LIBOR AND OVERNIGHT RATES

Most of our discussion so far has centered around Treasury rates. We now mention other rates which are, or have in the past, been important.

Libor used to be used as an important reference rate for trillions of dollars of transactions. As a result of scandals surrounding Libor, which are discussed in Chapter 16 of FMP, it has been replaced by rates calculated by compounding overnight reference rates. In the United States, the overnight reference rate is the Secured Overnight Financing Rate (SOFR). This is an average of the rates on overnight repo transactions. In other countries, the overnight reference rates are calculated from unsecured borrowing and lending between banks at the end of each day. Banks with surplus funds at the end of a day lend overnight to banks that need funds. A broker usually matches borrowers and lenders. In the United Kingdom, the overnight reference rate, calculated in this way, is the sterling overnight index average (SONIA), while in the Eurozone it is the euro overnight index average (EONIA).

10.8 SWAPS

Swaps are discussed in some detail in Chapter 20 of Financial Markets and Products. The floating rate in the United States that has replace LIBOR is SOFR. Party A could agree to pay Party B interest at a fixed rate of 4% per annum on a notional principal of USD 100 million. In return Party B might agree to pay three month SOFR, quarter by quarter.

A possible outcome of this transaction is shown in Table 10.6. It is assumed that three-month SOFR is 3% for the first three months. This means that USD 0.75 (= $0.25 \times 0.03 \times 100$) million is paid by Party B to Party A at time 0.25 years. Three month

Table 10.6 A Possible Outcome for the Swap Considered in the Text

Time	SOFR rate for previous 3 months	Party B payment	Party A Payment	Net Payment Party B to Party A
0.25	3%	0.75	1.0	−0.25
0.50	3.2%	0.80	1.0	−0.20
0.75	3.4%	0.85	1.0	−0.15
...				
...				
4.75	5.4%	1.35	1.0	0.35
5.00	5.6%	1.40	1.0	0.40

SOFR is 3.2% during the second three months. This leads to a payment of USD 0.80 ($= 0.25 \times 0.032 \times 100$) million at time 0.5 years by Party B to Party A. A fixed payment of USD 1.0 ($= 0.25 \times 0.04 \times 100$) million is made by Party A to Party B on all 20 payment days.

The principal (USD 100 million in this example is referred to as the notional principal because it is never actually exchanged. However, the principal could be exchanged at the end of the five years. Party A would pay Party B USD 100 million and Party B would pay Party A USD 100 million. This exchange has no value. The swap can therefore be considered as exchanging a fixed rate bond for a floating rate bond. The floating rate bond is worth par. The same should therefore be true of the fixed-rate bond at the time the swap is initiated. The swap market therefore defines a series of par rate bonds. (See Section 10.3 for a discussion of par rates.)

SUMMARY

The compounding frequency used for an interest rate defines the unit of measurement. If an interest rate is measured with annual compounding, the quoted interest rate is assumed to be compounded once a year; if measured with semi-annual compounding, it is assumed to be compounded twice a year; and so on. Interest rates in derivatives markets are usually measured with continuous compounding.

The spot rate for a certain maturity is the zero-coupon interest rate for that maturity. It can be calculated directly from the discount factor for that maturity. The forward rate for a future period is the spot rate for that period implicit in the spot rates observed in the market today. The par rate is the coupon rate on a bond that causes the bond price to equal par.

When the term structure is upward-sloping, the forward rate for a period beginning at time T is greater than the T-year spot rate which in turn is greater than the T-year par rate. When the term structure is downward-sloping, the forward rate for a period beginning at time T is less than the T-year spot rate, which in turn is less than the T-year par rate.

Swap rates are the fixed rates exchanged for floating rates in a swap agreement. Swap rates are par rates. This means that the swap rates observed in the market at a particular time can be used to define bonds that sell for par. These can in turn be used to calculate discount factors and spot rates.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

QUESTIONS

Short Concept Questions

10.1 If you are receiving funds at 5% interest, would you prefer the rate to be measured with semi-annual compounding or quarterly compounding?

10.2 What rate with annual compounding is equivalent to 10% with semi-annual compounding?

10.3 What rate with continuous compounding is equivalent to 7% with annual compounding?

10.4 What is another name for the spot rate?

10.5 If the two-year rate and the three-year rate are 4% and 3%, respectively, what is the forward rate for the third year? Assume all rates are measured with continuous compounding.

10.6 In an upward-sloping yield curve environment, which is higher: the four-year par yield or the four-year spot rate?

10.7 In a downward-sloping yield curve environment, which is higher: the five-year spot rate or the forward rate for the period between five and six years?

10.8 How is a flattening of the term structure defined?

10.9 What floating rate has replaced Libor in the United States? How is it calculated?

10.10 What trade should you put on if you feel that the actual interest rate for the third year will be less than the forward rate for the third year?

Practice Questions

10.11 An investor pays USD 100 and receives USD 109 in one year. What is the interest rate with semi-annual compounding?

10.12 What is the present value of USD 100 received in eight years when the spot rate is 5% per annum with semi-annual compounding? What is the eight-year discount factor?

10.13 A rate is 5.5% with continuous compounding. What is the equivalent rate with quarterly compounding?

10.14 Spot rates with semi-annual compounding are as follows.

Maturity	Rate
0.5	3.0
1.0	3.5
1.5	3.8
2.0	4.0
2.5	4.1
3.0	4.2

Calculate forward rates for each six-month period with semi-annual compounding.

10.15 If spot rates are as in Question 10.14, what are par rates for one, two, and three years with semi-annual compounding?

10.16 An annuity that pays USD 1 on each coupon payment date of a five-year Treasury bond is worth USD 8.94. The five-year par rate is 4%. What is the value of the bond if it pays a coupon of 6%?

10.17 The coupon rate on a five-year bond is higher than the forward rate between time 4.5 years and time five years. If forward rates do not change do you expect the bond price to increase or decrease during the next six months?

10.18 What trading strategy should be followed if it is considered that the term structure will steepen?

10.19 One-, two-, and three-year swap rates where payments are exchanged semi-annually are 3%, 3.4%, and 3.7%. Explain how you would use this data to determine one-, two-, and three-year spot rates.

10.20 The cash prices of 6-month and one-year Treasury bills are 97.0 and 93.0. A 1.5-year and two-year Treasury bond with coupons at the rate of 6% per year sell for 98.5 and 97.5. Calculate the six-month, 12-month, 18-month, and 24-month spot rates with semi-annual compounding.

ANSWERS

Short Concept Questions

10.1 You would like the rate to be measured with quarterly compounding (the higher the compounding frequency the higher the rate).

10.2 The rate is $1.05^2 - 1 = 0.1025$ or 10.25%

10.3 The rate is $\ln(1.07) = 0.0677$ or 6.77%

10.4 Zero-coupon interest rate or simply zero.

10.5 The forward rate is 1%.

10.6 The four-year spot rate is higher.

10.7 The five-year spot rate is higher.

10.8 A flattening of the yield curve occurs when long-maturity rates move down by more than short-maturity rates or when long-maturity rates move up by less than short-maturity rates.

10.9 SOFR. This is calculated by compounding overnight rates.

10.10 You should borrow for two years and invest for three years. If you are right, borrowing for two years then taking an extra loan for the third year will be less expensive than borrowing for three years.

Solved Problems

10.11 8.806%

10.12 $\dfrac{100}{1.025^{16}} = 67.36$

10.13 $4(e^{0.055/4} - 1) = 0.05538$. The equivalent rate is 5.538%.

10.14 The forward rate per six months for the period between 0.5 and one year is, from Equation (10.5):

$$\frac{(1 + 0.035/2)^2}{(1 + 0.03/2)} - 1 = 0.0200062$$

This is 4.0012% per annum with semi-annual compounding. The forward rate per six months for the period between one and 1.5 years is

$$\frac{(1 + 0.038/2)^3}{(1 + 0.035/2)^2} - 1 = 0.0220066$$

This is 4.4013% per year with semi-annual compounding. Similarly the forward rates for six month periods starting in 1.5, 2, 2.5, and 3 years are 4.6012%, 4.5005%, and 4.7007%

10.15 The one-year par rate (from Equation 10.4) is

$$\frac{2 \times 100 \times (1 - 0.965898)}{(0.985222 + 0.965898)} = 3.4957$$

Similarly, the two and three-year par rates are 3.9871% and 4.1822%.

10.16 The value is

$$100 + \frac{6 - 4}{2} \times 8.94 = 108.94$$

10.17 It will decrease. The decrease in value is the value of a lost forward rate agreement, which in this case is positive.

10.18 A trader should sell long-maturity bonds and buy short-maturity bonds.

10.19 The three swap rates define par yield bonds. One, two, and three-year bonds, which pay coupons semi-annually at the rate of 3%, 3.4%, and 3.7% selling for par. These can be used to imply spot rates.

10.20 The six-month rate (semi-annually compounded) is

$$2\left(\frac{100}{97} - 1\right) = 0.06186$$

or 6.186%. The one-year rate (semi-annually compounded) is

$$2 \times \left[\left(\frac{100}{93}\right)^{1/2} - 1\right] = 0.07390$$

or 7.390%.

The coupons on the 1.5 year bond have a value of $0.97 \times 3 + 0.93 \times 3 = 5.7$. The value of the final payment is therefore $98.5 - 5.7 = 92.8$. The discount factor for 1.5 years is $92.8/103 = 0.900971$. This corresponds to a spot rate (semi-annually compounded) of 7.074%.

The coupons on the two-year bond have a value of

$$3 \times 0.97 + 3 \times 0.93 + 3 \times 0.900971 = 8.4029$$

The value of the final payment is therefore $97.5 - 8.4029 = 89.0971$. The discount factor for two years is $89.0971/103 = 0.8650$. This corresponds to a spot rate (semi-annually compounded) of 7.383%.

11

Bond Yields and Return Calculations

Learning Objectives

After completing this reading, you should be able to:

- Differentiate between gross and net realized returns and calculate the realized return for a bond over a holding period including reinvestments.
- Define and interpret the spread of a bond and explain how a spread is derived from a bond price and a term structure of rates.
- Define, interpret, and apply a bond's yield to maturity (YTM) to bond pricing.
- Explain how to calculate a bond's YTM given its structure and price.
- Calculate the price of an annuity and a perpetuity.
- Explain the relationship between spot rates and YTM.
- Define the coupon effect and explain the relationship between coupon rate, YTM, and bond prices.
- Explain the decomposition of the profit and loss (P&L) for a bond position or portfolio into separate factors including carry roll-down, rate change, and spread change effects.
- Describe the common assumptions made about interest rates when calculating carry roll-down, and calculate carry roll-down under these assumptions.

In this chapter, we continue to discuss the terminology used in fixed-income markets. We explain how bond yields are defined. We also describe how returns from trading strategies are calculated (both on a gross and a net basis) and how the spread implicit in a trading strategy can be derived.

We also discuss how a trading strategy's profit and loss can be broken down into its component parts: the carry roll-down, an amount due to rate changes, and an amount due to spread changes.

11.1 REALIZED RETURN

A bond's realized return is calculated by comparing the initial investment's value with its final value. To take a simple example, suppose a bond is bought for USD 98 immediately after a coupon payment date. It earns a coupon of USD 1.75 in six months and is worth USD 98.5 at that time. The capital gain from the increase in the bond's price is USD 0.5 (= 98.5 − 98.0). The realized return from the bond over six months is therefore:

$$\frac{98.5 - 98.0 + 1.75}{98.0} = 0.02296$$

This is 4.592% (= 2 × 2.296%) per annum with semi-annual compounding.[1]

If we want to look at the return over a longer period (e.g., one year), we must consider the investment of the coupon received at the six-month point as well. Suppose the invested proceeds from the coupon earn 1.1% for the following six months and that the bond is worth USD 98.7 after one year. The realized return on the bond over the one-year period is

$$\frac{98.7 - 98.0 + 1.75 + 1.75 \times 1.011}{98.0} = 0.04305$$

or 4.305% per annum with annual compounding (equivalent to 4.26% with semi-annual compounding). Returns over longer periods can be calculated similarly. These returns are referred to as *gross returns* because the cost of financing the bond purchase is not considered.

The *net return* is the return after financing costs have been subtracted. Suppose that the position in the bond has been financed at 3% per annum (semi-annually compounded). The cost of financing the purchase of the bond for six months in our first example is USD 1.47 (= 98 × 0.015). The net realized return over six months is therefore:

$$\frac{98.5 - 98.0 + 1.75 - 1.47}{98.0} = 0.00796$$

[1] The annual equivalent rate would be $(1.02296^2) - 1 = 4.64\%$

This is 1.592% (= 2 × 0.796%) per annum with semi-annual compounding.

The cost of financing the bond for one year is USD 2.962 (= $98 \times (1.015^2 - 1)$). The net realized return over one year is therefore:

$$\frac{98.7 - 98.0 + 1.75 + 1.75 \times 1.011 - 2.962}{98.0} = 0.01283$$

or 1.283% (with annual compounding).

This is the usual method of calculating the net return. An alternative approach would be to compare the profit to the net outlay. When the bond is fully financed, however, the net outlay is zero and a meaningful net return cannot be calculated. (The return calculated would be infinite.)

11.2 SPREADS

An investor buying a Treasury security may be interested in calculating the excess return earned over the return provided by other Treasury securities. This can be done by asking the following question:

> *What spread do we have to add to the prevailing Treasury forward rates so that the present value of the cash flows on the security equals the price paid for the security?*

Suppose Treasury forward rates (expressed with semi-annual compounding) are as presented in Table 11.1.

As pointed out in Equation (10.5) of Chapter 10, spot rates can be calculated by compounding forward rates. Discount factors (covered in Chapter 9) can be calculated in a direct way from spot rates. The six-month discount factor in Table 11.1 is

$$\frac{1}{1 + 0.007/2} = 0.996512$$

The 12-month discount factor is

$$\frac{1}{(1 + 0.007/2) \times (1 + 0.012/2)} = 0.990569$$

Table 11.1 Treasury Forward Rates with Semi-Annual Compounding

Period (Months)	Forward Rate (%)
0–6	0.7
6–12	1.2
12–18	1.6
18–24	2.0

The 18-month discount factor is

$$\frac{1}{(1 + 0.007/2) \times (1 + 0.012/2) \times (1 + 0.016/2)} = 0.982707$$

Finally, the 24-month discount factor is

$$\frac{1}{(1 + 0.007/2) \times (1 + 0.012/2) \times (1 + 0.016/2) \times (1 + 0.020/2)} = 0.972977$$

These discount factors are summarized in Table 11.2.

Now suppose that an investor buys a two-year Treasury bond with a coupon of 2.5% for USD 101.5. The first thing the investor might do is value the bond. The value of the bond (per USD 100 of face value) from Table 11.2 is

$$1.25 \times 0.996512 + 1.25 \times 0.990569 + 1.25 \times 0.982707 + 101.25 \times 0.972977 = 102.226$$

The investor has managed to buy the bond for USD 101.5 when the theoretical value is USD 102.226. This is a gain of USD 0.726. To convert this gain into a spread, we consider how much the forward rates would have to be increased for the theoretical price to be USD 101.5. Suppose we add a spread of s to each forward rate. The discount factors for 6 months, 12 months, 18 months, and 24 months become

$$\frac{1}{1 + 0.007/2 + s/2}$$

$$\frac{1}{(1 + 0.007/2 + s/2)(1 + 0.012/2 + s/2)}$$

$$\frac{1}{(1 + 0.007/2 + s/2)(1 + 0.012/2 + s/2)(1 + 0.016/2 + s/2)}$$

and

$$\frac{1}{(1 + 0.007/2 + s/2)(1 + 0.012/2 + s/2)(1 + 0.016/2 + s/2)(1 + 0.02/2 + s/2)}$$

Thus, the value of the required spread is determined by solving

$$\frac{1.25}{1 + 0.0035 + 0.5s} + \frac{1.25}{(1 + 0.0035 + 0.5s)(1 + 0.006 + 0.5s)} + \frac{1.25}{(1 + 0.0035 + 0.5s)(1 + 0.006 + 0.5s)(1 + 0.008 + 0.5s)}$$
$$+ \frac{101.25}{(1 + 0.0035 + 0.5s)(1 + 0.006 + 0.5s)(1 + 0.008 + 0.5s)(1 + .01 + 0.5s)} = 101.5$$

The solution (which can be found by trial and error) is $s = 0.00366$ (or 0.366%).[2] The spread earned is therefore 36.6 basis points.

In this case, we have calculated the spread earned by buying a Treasury bond at a favorable price (relative to other Treasury bonds in the market). Another type of calculation sometimes carried out compares one market to another market.

For example, we might be interested in the spread between the AA-rated corporate bond market and the Treasury market. We would then calculate the spread, which when added to Treasury forward rates, gives the market price of an AA-rated bond. In general, the spread calculated will depend on maturity. For example, we might find three-year AA-rated bonds provide a spread over Treasuries of 50 basis points per year, whereas five-year AA-rated bonds provide a spread over Treasuries of 80 basis points per year.

Table 11.2 Discount Factors Calculated from Table 11.1

Maturity (Years)	Discount Factor
0.5	0.996512
1.0	0.990569
1.5	0.982707
2.0	0.972977

11.3 YIELD TO MATURITY

A bond's yield to maturity is the single discount rate, which if applied to all the bond's cash flows, would make the cash flows' present value equal to the bond's market price. For example, suppose a two-year bond with a coupon of 2.5% sells for USD 102. The yield y (expressed with semi-annual compounding) is the solution to:

$$102 = \frac{1.25}{1 + y/2} + \frac{1.25}{(1 + y/2)^2} + \frac{1.25}{(1 + y/2)^3} + \frac{101.25}{(1 + y/2)^4}$$

This can be solved by trial and error to give $y = 1.48\%$. To check that this is correct we note

$$102 = \frac{1.25}{1 + 0.0074} + \frac{1.25}{(1 + 0.0074)^2} + \frac{1.25}{(1 + 0.0074)^3} + \frac{101.25}{(1 + 0.0074)^4}$$

The yield to maturity is a convenient measure because the price of a bond can be unambiguously converted into its yield to maturity (and vice versa).

[2] Solver in Excel is a useful resource for solving this type of non-linear equation.

Formally, the yield to maturity (expressed with semi-annual compounding) for a bond with price P, lasting T years, and paying a coupon of c, is the value of y that solves

$$P = \frac{c/2}{1 + y/2} + \frac{c/2}{(1 + y/2)^2} + \frac{c/2}{(1 + y/2)^3} + \cdots + \frac{100 + c/2}{(1 + y/2)^{2T}}$$
$$= \frac{c}{2}\sum_{i=1}^{2T}\left(\frac{1}{1 + y/2}\right)^i + \frac{100}{(1 + y/2)^{2T}} \quad (11.1)$$

This equation assumes the price is observed immediately after a coupon payment date (i.e., that T is an integral number of half years). If the bond is being observed between coupon payment dates, and P is the quoted price, we must first calculate the cash (dirty) price. This (as explained in Chapter 9) is the quoted price plus accrued interest. It is the price that would be paid by the purchaser (and received by the seller).

Suppose that the dirty price is P^* and that coupons are received at times $t_1, t_2, \ldots, t_n$ with $t_n = T$. Equation (11.1) becomes

$$P^* = \frac{c/2}{(1 + y/2)^{2t_1}} + \frac{c/2}{(1 + y/2)^{2t_2}} + \frac{c/2}{(1 + y/2)^{2t_3}}$$
$$+ \cdots + \frac{c/2}{(1 + y/2)^{2t_{n-1}}} + \frac{100 + c/2}{(1 + y/2)^{2T}}$$
$$= \frac{c}{2}\sum_{i=1}^{n}\left(\frac{1}{1 + y/2}\right)^{2t_i} + \frac{100}{(1 + y/2)^{2T}}$$

Because coupons are paid every six months, $t_i = t_1 + 0.5(i-1)$ and P^* can be written as:

$$P^* = \frac{c}{2}\left(\frac{1}{1 + y/2}\right)^{2t_1} \times \sum_{i=0}^{n-1}\left(\frac{1}{1 + y/2}\right)^i + \frac{100}{(1 + y/2)^{2T}} \quad (11.2)$$

In this more general situation, $T = t_1 + 0.5(n - 1)$ so that $2T = 2t_1 + n - 1$.

As an example, suppose coupon payments on a bond are made at times 0.2, 0.7, 1.2, 1.7, and 2.2 at the rate of 5% per year. The dirty price of the bond is USD 97. To find the yield, we must solve

$$97 = \frac{2.5}{(1 + y/2)^{2\times0.2}} + \frac{2.5}{(1 + y/2)^{2\times0.7}}$$
$$+ \frac{2.5}{(1 + y/2)^{2\times1.2}} + \frac{2.5}{(1 + y/2)^{2\times1.7}} + \frac{102.5}{(1 + y/2)^{2\times2.2}}$$
$$= 2.5\left(\frac{1}{(1 + y/2)^{2\times0.2}}\right)\sum_{i=0}^{4}\frac{1}{(1 + y/2)^i} + \frac{100}{(1 + y/2)^{2T}}$$

The solution to this is $y = 7.24\%$.

Annuities

From the mathematics of geometric series, we know that:

$$x + x^2 + x^3 \cdots + x^n = \frac{x(1 - x^n)}{1 - x}$$

For example, setting $x = 1/2$ and $n = 3$ gives

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{(1/2)(1 - 0.5^3)}{1 - 0.5} = 0.875$$

We first use this to simplify Equation (11.1) by setting

$$x = \frac{1}{1 + y/2}$$

This means that:

$$\sum_{i=1}^{2T}\left(\frac{1}{1 + y/2}\right)^i = \frac{1/(1 + y/2)}{1 - 1/(1 + y/2)}\left[1 - \frac{1}{(1 + y/2)^{2T}}\right]$$
$$= \frac{2}{y}\left[1 - \frac{1}{(1 + y/2)^{2T}}\right]$$

The value of an annuity paying c/2 every six months is c/2 multiplied by this, or:

$$\frac{c}{y}\left[1 - \frac{1}{(1+y/2)^{2T}}\right] \quad (11.3)$$

Equation (11.1) can be simplified to:

$$P = \frac{c}{y}\left[1 - \frac{1}{(1 + y/2)^{2T}}\right] + \frac{100}{(1 + y/2)^{2T}}$$

Equation (11.2) can similarly be simplified to:

$$P^* = \frac{c}{y}\left(\frac{1}{1 + y/2}\right)^{2t_1 - 1}\left[1 - \frac{1}{(1 + y/2)^n}\right] + \frac{100}{(1 + y/2)^{2T}}$$

Perpetuity

In a perpetuity, payments are received semi-annually at the rate of c per year and last forever. The value of a perpetuity is given by letting T tend to infinity in Equation (11.3). When we do this, the term:

$$\frac{1}{(1 + y/2)^{2T}}$$

disappears. Thus, the perpetuity's value can be seen to be

$$\frac{c}{y}$$

Suppose we receive a semi-annual coupon at the rate of USD 5 per annum for ten years. Equation (11.3) shows that when the yield to maturity is 8%, this is worth:

$$\frac{5}{0.08}\left[1 - \frac{1}{1.04^{20}}\right] = 33.98$$

If this annuity becomes a perpetuity, its value becomes USD 62.50 (= 5/.08).

Properties of Yield to Maturity

Some straightforward properties of the yield to maturity are as follows.

- When the yield to maturity is equal to the coupon rate, the bond sells for its face value.
- When the yield to maturity is less than the coupon rate, the bond sells for more than its face value. If time passes with no change to the yield, the price of the bond declines.
- When the yield to maturity is greater than the coupon rate, the bond sells for less than its face value. If time passes with no change to the yield to maturity, the price of the bond increases.
- If the term structure is flat with all rates equal to *R*, the yield to maturity is equal to *R* for all maturities.

The Coupon Effect

Bonds with the same maturity and different coupons do not necessarily have the same yield to maturity. To illustrate this, consider the two term structures in Table 11.3. Term structure A is upward-sloping, whereas term structure B is downward-sloping. Table 11.4 calculates the yield to maturity given term structure A for bonds with coupons of 0%, 2%, 4%, 6%, 8%, and 10%. Not surprisingly, the yield to maturity when the coupon is 0% is exactly equal to the spot rate for five years (or 5.2%). The yield to maturity decreases as the coupon increases. Table 11.5 calculates the yield to maturity given (the downward-sloping) term structure B for bonds with the same coupons. Again, we find that the yield to maturity when the coupon is 0% is the five-year rate. But in this case the yield increases as the coupon rate increases.

Table 11.3 Upward- and Downward-Sloping Term Structures. Spot Rates Expressed with Semi-Annual Compounding

Maturity (Years)	Spot Rate: Term Structure A (%)	Spot Rate: Term Structure B (%)
0.5	2.00	5.05
1.0	3.00	4.65
1.5	3.60	4.35
2.0	4.00	4.10
2.5	4.40	3.90
3.0	4.65	3.75
3.5	4.85	3.65
4.0	5.00	3.55
4.5	5.10	3.50
5.0	5.20	3.45

Table 11.4 Variation of Yield to Maturity of a Five-Year Bond with Coupon for Upward-Sloping Term Structure A in Table 11.3, with Bond Yields Expressed with Semi-Annual Compounding

Bond Coupon	Bond Price (USD)	Bond Yield to Maturity (%)
0%	77.36	5.200
2%	86.20	5.167
4%	95.04	5.138
6%	103.88	5.112
8%	112.71	5.088
10%	121.55	5.067

To understand these results, note that the yield to maturity is a complicated function of the spot rates applicable to each payment date. As the coupon increases, the average time until the bond holder receives cash flows decreases, and the spot rates for the early payment dates become relatively more important in determining the yield. For an upward-sloping interest rate term structure (such as that in Table 11.4), the spot rates for the early payment dates are lower than the spot rate for the final payment date. As a result, the yield to maturity declines as the coupon rate increases. Similarly, in a downward-sloping term structure environment (such as that in Table 11.5), the yield to maturity increases as the coupon rate increases.

The fact that correctly priced bonds with the same maturity and different coupons have different yields to maturity is called the *coupon effect*. As Table 11.4 illustrates, yield is not a reliable measure of relative value. The bond in Table 11.4 with a 4% coupon has a lower yield than the bond with a 2% coupon. However, it is worth more.

Table 11.5 Variation of Yield to Maturity of a Five-Year Bond with Coupon for Downward-Sloping Term Structure B in Table 11.3 with Bond Yields Expressed with Semi-Annual Compounding

Bond Coupon	Bond Price (USD)	Bond Yield to Maturity (%)
0%	84.28	3.450
2%	93.32	3.467
4%	102.36	3.481
6%	111.40	3.495
8%	120.44	3.507
10%	129.48	3.518

Japanese Yields

The conventions for quoting rates and yields vary from country to country. In the last three chapters, we have focused on the conventions used for Treasury bonds in the United States. Note that in Japan, yields are quoted on a *simple yield* basis. This means that there is no compounding in the yield measurement.

Consider a bond with a face value of JPY 100 and T years to maturity whose price is p and pays a coupon of c. The coupon provides a (simple interest) return of c/p. The difference between the price and the face value provides a further (simple interest) return of $(100-p)/p$ over T years. This is

$$\frac{1}{T}\frac{100-p}{p}$$

per year. Thus, the yield is

$$y = \frac{c}{p} + \frac{100-p}{pT}$$

Suppose, for example, that a five-year bond has a coupon of 2% and the price of the bond is JPY 99. The yield is

$$\frac{2}{99} + \frac{100-99}{99 \times 5} = 0.0222$$

or 2.22%.[3]

11.4 CARRY ROLL-DOWN

The carry roll-down is designed to estimate the return achieved if there is no change to some aspect of the interest rate environment.

The most common assumption when the carry roll-down is calculated is that forward rates are realized (i.e., the forward rate for a future period remains unchanged as we move through time). When the beginning of the future period is reached, the spot rate for the period equals the forward rate.

For example, suppose the term structure is flat at 4% (with semi-annual compounding) and that an investor owns a five-year bond paying a 4% coupon with a face value of USD 100. The price of the bond is USD 100, and all forward rates are 4%. For the carry roll-down, we assume forward rates remain unchanged as we move through time. This means that the term structure continues to be flat at 4%. The value of the bond will then stay at USD 100. The investor then earns a coupon of 2% and has no price appreciation or depreciation. The carry roll-down is therefore USD 2.00 for the six-month period.

[3] For more information on the yield calculation for Japanese bonds, see the Japan Ministry of Finance publication: https://www.mof.go.jp/english/jgbs/debt_management/guide.htm

Table 11.6 Forward Rates for a Two-Year Bond

Period (Months)	Forward Rate (%) (Semi-Annually Compounded)
0–6	0.7
6–12	1.2
12–18	1.6
18–24	2.0

Now consider a slightly more complicated example. The term structure is still flat at 4% (with semi-annual compounding) and an investor owns a five-year bond paying a 5% coupon. In this case, the price of the bond is USD 104.49. If the term structure stays flat at 4%, the bond will have a maturity of 4.5 years at the end of six months and will be worth USD 104.08. (The bond is providing a relatively high coupon, but as time passes the coupon will be received for a shorter period of time and so the bond price declines.)

The value of the bond declines by USD 0.41 (= 104.49 − 104.08) and the carry roll-down is therefore USD 2.09 (= 2.50 − 0.41). The USD 2.50 is the coupon received in cash and is sometimes referred to as the *cash-carry*. The −0.41 is the price-change component of the carry roll-down.

For a more complete example, consider again a Treasury bond with a coupon of 2.5%. The forward rates for this bond are shown in Table 11.6 and the bond is currently valued at USD 102.226. To calculate the carry roll-down, we assume that the forward rates are realized. This means that after six months, the forward rates for the 0–6, 6–12, and 12–18 month periods are the same as those observed today for the 6–12, 12–18, and 18–24 month periods. This assumption is shown in Table 11.7.

The price of the bond at the end of six months under the carry roll-down assumption is

$$\frac{1.25}{1.006} + \frac{1.25}{1.006 \times 1.008} + \frac{101.25}{1.006 \times 1.008 \times 1.01} = 101.334$$

Table 11.7 Forward Rates After Six Months for Example in Table 11.6 if Forward Rates Are Realized

Period (Months)	Forward Rate (%) (Semi-Annually Compounded)
0–6	1.2
6–12	1.6
12–18	2.0

The calculations show that if forward rates are realized, we expect the price of the bond to decrease from USD 102.226 to USD 101.334 after six months. During this time, a coupon of USD 1.25 will be received. The carry roll-down per USD 100 face value is therefore USD 0.358 (= 101.334 − 102.226 + 1.25).

A quicker way of calculating the carry roll-down (assuming forward rates are realized) is to assume the return earned on any bond over the next period is always the prevailing one-period rate. In our example, the six-month rate is 0.7% semi-annually compounded (see Table 11.6). The return earned over the next six months is therefore 0.35%. The current value of the bond is USD 102.226. This carry roll-down is therefore:

$$0.0035 \times 102.226 = 0.358$$

This agrees with our earlier calculation.

Note that the bond portfolio we are considering here does not matter. If forward rates are realized over the next period, the portfolio will always earn the prevailing market rate for the next period. The carry roll-down for the portfolio is the prevailing rate for the next period applied to the current value of the portfolio. We can extend this result so that it applies to several periods. If forward rates are realized for several periods, the gross return realized will be the rate in the market applicable to those periods.

This result has implications for trading strategies. Should you buy long-maturity or short-maturity bonds? If you expect forward rates to be realized, you will get the same return in both cases. If you expect realized rates to be less than forward rates, however, long-maturity bonds will provide a better gross return. If you expect realized rates to be greater than forward rates, the reverse is true, and a series of short maturity bonds will provide the best gross return.

Alternative Carry Roll-Down Assumptions

An alternative to the "forward rates are realized" assumption in carry roll-down calculations is to assume the interest rate term structure stays unchanged. An unchanged term structure would mean that the forward rates in Table 11.6 become those in Table 11.8.

Table 11.8 Rates for Example in Table 11.1 After Six Months if Term Structure Remains Unchanged

Period (Months)	Forward Rate (%)
0–6	0.7
6–12	1.2
12–18	1.6

The bond price at the end of six months under the unchanged term structure assumption is

$$\frac{1.25}{1.0035} + \frac{1.25}{1.0035 \times 1.006} + \frac{101.25}{1.0035 \times 1.006 \times 1.008} = 101.983$$

The carry roll-down is then USD 1.007 (= 101.983 − 102.226 + 1.25) per USD 100 of face value.

An argument in favor of the unchanged term structure assumption is that an upward-sloping term structure reflects investor risk preferences. Investors demand an extra return to induce them to invest for long maturities. If investors' risk preferences are not expected to change, the term structure can be reasonably expected to retain its shape.

Another assumption sometimes made in carry roll-down calculations is that a bond's yield to maturity will remain unchanged. The one-period gross return, assuming the yield remains unchanged, is the yield itself. If the coupons are invested at the yield over many periods, the gross return is also the yield. A criticism of this assumption is that we do not normally expect the yield on a bond to remain unchanged. In an upward-sloping term structure environment, we expect the yield of a coupon-bearing bond to increase as we approach the bond's maturity. Similarly, in a downward-sloping term structure environment, we expect the yield to decrease as we approach maturity.

A final possibility is for an investor to make personal estimates of future rates and use these as the basis for calculating the carry roll-down.

11.5 P&L COMPONENTS

We now explain how the profit and loss (P&L) from a fixed-income portfolio can be split into several components. The components are as follows.

- *The Carry Roll-Down*: This has already been discussed.
- *Rate Changes*: This is the return realized when realized rates differ from those assumed in the carry roll-down.
- *Spread Changes*: This is the return realized when a bond's spread relative to other bonds changes.

Suppose the bond considered earlier that provides a coupon of 2.5% is purchased for USD 101.5. Forward rates are as in Table 11.6. We saw in Section 11.2 that the market price of the bond is USD 102.226 and that, by buying the bond for USD 101.5, the investor earns a spread of 36.6 basis points per year.

We assume the carry roll-down is calculated assuming that forward rates are realized. We also assume the spread is unchanged. This means that the value of the bond (USD) at the end of six months is

$$\frac{1.25}{1+0.006+s/2}+\frac{1.25}{(1+0.006+s/2)(1+0.008+s/2)}$$
$$+\frac{101.25}{(1+0.006+s/2)(1+0.008+s/2)(1+0.01+s/2)}$$

with $s = 0.00366$. The value is USD 100.79, the cash-carry is USD 1.25. and the carry roll-down (USD) is therefore:

$$100.79 - 101.5 + 1.25 = 0.54$$

This can also be calculated by applying the six-month return of 0.7% per year, plus the spread of 0.366% per year, to the investment of USD 101.5:

Carry roll-down (USD) $= 0.5 \times (0.007 + 0.00366) \times 101.5 = 0.54$

Suppose now that forward rates are not realized. Instead, assume the forward rates in six months are those shown in Table 11.9. We continue to assume a spread of 36.6 basis points and calculate the value of the bonds as:

$$\frac{1.25}{1+0.005+s/2}+\frac{1.25}{(1+0.005+s/2)(1+0.007+s/2)}$$
$$+\frac{101.25}{(1+0.005+s/2)(1+0.007+s/2)(1+0.009+s/2)}$$

with $s = 0.00366$. The value is USD 101.09. If the spread had remained the same, the impact of the term structure change shown in Table 11.9 would therefore have been USD 0.30 (= 101.09 − 100.79).

Finally, suppose the spread does not stay at 36.6 basis points and instead drops to 30 basis points. The value of the bond is now given by:

$$\frac{1.25}{1+0.005+s/2}+\frac{1.25}{(1+0.005+s/2)(1+0.007+s/2)}$$
$$+\frac{101.25}{(1+0.005+s/2)(1+0.007+s/2)(1+0.009+s/2)}$$

with $s = 0.0030$ instead of $s = 0.00366$. The price is then USD 101.19, and the impact of the spread change is therefore USD 0.10 (= 101.19 − 101.09).

Table 11.9 Actual Term Structure in Six Months

Period (Months)	Forward Rate (%)
0–6	1.0
6–12	1.4
12–18	1.8

Table 11.10 Summary of P&L Decomposition

Initial Price of Bond (USD)	101.50
Carry Roll-Down (USD)	0.54
Rate Changes (USD)	0.30
Spread Changes (USD)	0.10
Final Value of Bond (USD)	101.19
Cash-Carry (USD)	1.25

These results are summarized in Table 11.10.

Note that the gain (USD) is

$$101.19 + 1.25 - 101.5 = 0.94$$

The P&L decomposition thus splits this into:

(a) A carry roll-down of USD 0.54,

(b) The impact of a term structure change of USD 0.30, and

(c) A spread change of USD 0.10.

The sum of the components equals the total gain:

$$0.94 = 0.54 + 0.30 + 0.10$$

If each component of the P&L is divided by the price paid for the bond, we obtain components of the gross return. In our example; the total gross return is 0.93%, (= 0.94/101.5); the carry roll-down return is 0.53% (= 0.54/101.5); the return due to rate changes is 0.30% (= 0.30/101.5); and the return due to the spread change is 0.10% (= 0.10/101.5).

Extensions

Two extensions of this analysis are worth mentioning. First, if there are financing costs, the analysis can be used to give components of the net return (or net P&L) rather than the gross return (or gross P&L). It is necessary to add a fourth (negative) component to the analysis, which is the cost of financing.

For ease of exposition, we have assumed we are considering the return between two coupon payment dates. In the more general situation, both the initial date and the final date are between coupon payment dates. We can then split the change in the value of the position of the bond into:

(a) A carry roll-down,

(b) The impact of rate changes,

(c) The impact of spread change, and

(d) The impact of accrued interest.

The quantities (a), (b), and (c) are calculated in terms of quoted prices. The impact of accrued interest on the change in the value of the position is the accrued interest at the end of the period minus the accrued interest at the beginning of the period.[4]

SUMMARY

The gross return from trading bonds is the return as a percentage of the initial investment without considering financing costs. The net return subtracts financing costs from the gross return before dividing by the initial investment.

When an investor buys a bond at a favorable price, the spread is the amount that must be added to each forward rate to equate the value of the bond's cash flows to the price paid.

Yield to maturity is a popular measure for a bond's return. It has the advantage in that there is a one-to-one correspondence between the price and the yield to maturity. When the price is known, the yield to maturity can be calculated (and vice versa). A bond's yield to maturity depends on both its coupon and its time to maturity. Two bonds with the same time to maturity and different coupons generally have different yields. In an upward-sloping term structure environment, the yield decreases as the coupon increases. In a downward-sloping term structure, the yield increases as the coupon increases.

The profit or loss from a trading strategy can be decomposed into the carry roll-down, the amount resulting from interest rate changes, and the amount resulting from spread changes. The carry roll-down is usually defined as the impact of forward rates being realized (i.e., future forward rates being equal to today's forward rates). However, it can also be defined as the impact of the term structure remaining unchanged, or the impact of bond yields remaining unchanged. Spread changes arise from a bond's market price moving closer to (or further away from) its theoretical price.

[4] If there is no coupon payment, the carry roll-down is the change in the quoted price. If there is a coupon payment, the interest on the coupon payment between the time of the payment and the end of the period should in theory be considered. However, this is a second order effect and often ignored.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

QUESTIONS

Short Concept Questions

11.1 How are gross return and net return defined?

11.2 How would a trader of Treasury securities define the spread from trading an AAA-rated bond?

11.3 How is a bond's yield to maturity defined?

11.4 How is the yield to maturity on a bond affected by its coupon when the yield curve is (a) flat, (b) upward-sloping, and (c) downward-sloping?

11.5 If a coupon is less than the yield to maturity, is a bond's value greater than or less than its face value?

11.6 What is the yield to maturity of a stripped bond with maturity T?

11.7 In what way are Japanese bond yields calculated differently from U.S. Treasury bond yields?

11.8 What are the different assumptions that can be made when calculating the carry roll-down?

11.9 What are the components of the P&L decomposition discussed in this chapter?

11.10 If the yield on a bond decreases, does the bond's price increase or decrease?

Practice Questions

11.11 A bond paying a coupon at the rate of 6% is held for six months. The price at the beginning of the six months is USD 102 and the price at the end of the six months is 101. What is the gross return?

11.12 If the bond in Question 11.11 is financed at 2% (per annum), what is the net return?

11.13 The three-year spot rate is 4% (semi-annually compounded). An investor buys a three-year zero-coupon bond for USD 87.0. What is the spread?

11.14 The yield on a two-year bond that provides a coupon of 5% is 6% (semi-annually compounded). What is the price of the bond per USD 100 of face value?

11.15 Now suppose that the price of the bond in Question 11.14 is USD 98. What is the bond's yield to maturity?

11.16 A Japanese bond sells for JPY 95. It provides a coupon of 3% per year and lasts ten years. What yield would be calculated?

11.17 A bond is worth USD 103.00. The spot rate for the next six months is 5% per annum (semi-annually compounded). What is the carry roll-down using the "forward rates will be realized" assumption?

11.18 Under what circumstances is the carry roll-down the same for the following three assumptions: (a) forward rates are realized, (b) term structure is unchanged, and (c) yield to maturity is unchanged?

11.19 Describe the different calculations that must be carried out in a P&L decomposition.

11.20 The term structure is initially flat at 5%, and an investor buys a five-year bond with a face value of USD 100 and an annual coupon rate of 4%, paid semi-annually, at a spread of ten basis points. Carry out a P&L decomposition if, at the end of six months, the term structure is flat at 6% and there is no change in the spread.

ANSWERS

Short Concept Questions

11.1 Gross return is the return divided by the initial investment without considering financing costs. Net return is the return minus financing costs divided by the initial investment.

11.2 The trader would calculate the spread that has to be added to Treasury forward rates so that the resulting rates, when used for discounting, lead to the value of the AAA-rated bond bring equal to its market price.

11.3 A bond's yield to maturity is the single interest rate, which when used to discount all of the bond's cash flows, gives the market price of the bond.

11.4 When the yield curve is flat so that all rates equal R, the yield to maturity of a bond always equals R. When the yield curve is upward-sloping, the yield to maturity decreases as the coupon increases. When the yield curve is downward-sloping, the yield to maturity increases as the coupon increases.

11.5 The bond's value is less than the face value when the coupon is less than the yield to maturity.

11.6 The yield to maturity of a stripped bond with maturity T is the spot rate for maturity T.

11.7 The yield on Japanese bonds is calculated as a yield where there is no compounding. The yield is the annual coupon divided by the price plus the average return that will be provided per year as the price changes during the bond's life.

11.8 Carry roll-down can be calculated (a) assuming that forward rates are realized, (b) assuming that the term structure remains unchanged (c) assuming that bond yields remain unchanged, and (d) assuming that estimates made by the investor are realized.

11.9 The components of the P&L decomposition are (a) the carry roll-down, (b) the impact of rates being different from those assumed in the carry roll-down, and (c) the impact of spread changes. Financing costs and the changes in accrued interest could be additional components.

11.10 If the yield on a bond decreases, its price increases. Yields and prices are inversely related.

Solved Problems

11.11 The gross return is

$$\frac{3 + 101 - 102}{102} = 0.0196$$

or 1.96%.

11.12 If the bond is financed by 2% per annum, the net return for the six-month holding period is:

$$\frac{(3 + 101 - 102 - 0.01 \times 102)}{102} = 0.0096,$$

or 0.96%

11.13 Suppose that the spread is s. We need to solve

$$87 = \frac{100}{(1 + 0.04/2 + s/2)^6}$$

so that:

$$1 + 0.02 + s/2 = \left(\frac{100}{87}\right)^{1/6} = 1.0235$$

The spread is 0.0070 or 70 basis points.

11.14 The price of the bond is

$$\frac{2.5}{1.03} + \frac{2.5}{1.03^2} + \frac{2.5}{1.03^3} + \frac{102.5}{1.03^4} = 98.14$$

that is, USD 98.14 per USD 100 of face value.

11.15 If the price of the bond is USD 98, we must solve

$$\frac{2.5}{(1 + y/2)} + \frac{2.5}{(1 + y/2)^2} + \frac{2.5}{(1 + y/2)^3} + \frac{102.5}{(1 + y/2)^4} = 98$$

The solution (found by trial and error or using Excel's Solver) is 6.077%.

11.16 The yield is

$$\frac{3}{95} + \frac{1}{10}\frac{100 - 95}{95} = 0.0368$$

or 3.68%.

11.17 The carry roll-down is

$$0.025 \times 103 = 2.575$$

or USD 2.5756 per USD 100 of face value.

11.18 If the term structure is flat, the carry roll-down will be the same for all three definitions.

11.19 We first calculate the carry roll-down. This has two components: the cash-carry (i.e., the coupon payment) and the change in price. In calculating the carry roll-down, we assume that the spread remains unchanged. We then calculate the value of the bond for the actual change in the term structure assuming again that the spread remains unchanged. Finally, we calculate the effect on the bond price of the spread change.

11.20 First we calculate the carry roll-down. The cash-carry is 2%. In this case, the assumption underlying the carry roll-down is that the term structure remains flat at 5%. (This is true for all three definitions of carry roll-down.) The initial price paid for the bond is

$$2\sum_{i=1}^{10}\frac{1}{(1+0.025+0.001/2)^i}+\frac{100}{(1+0.025+0.001/2)^{10}}$$
$$=95.199$$

The price of the bond (USD), if six months passes without rates changing or the spread changing, is

$$2\sum_{i=1}^{9}\frac{1}{(1+0.025+0.001/2)^i}+\frac{100}{(1+0.025+0.001/2)^{9}}$$
$$=95.626$$

The carry roll-down (USD) is therefore:

$$2+95.626-95.199=2.427$$

This can also be calculated as 0.0255×95.199.

The value of the bond at the end of six months, assuming no spread change, is

$$2\sum_{i=1}^{9}\frac{1}{(1+0.03+0.001/2)^i}+\frac{100}{(1+0.03+0.001/2)^{9}}$$
$$=91.844$$

After the spread change is considered, the value of the bond is

$$2\sum_{i=1}^{9}\frac{1}{(1+0.03)^i}+\frac{100}{(1+0.03)^{9}}=92.214$$

This leads to the following table

Initial Bond Price (USD)	95.199
Carry Roll-Down (USD)	2.427
Impact of Term Structure Change (USD)	$91.844 - 95.626 = -3.782$
Impact of Spread Change (USD)	$92.214 - 91.844 = 0.370$
Final Bond Price (USD)	92.214
Cash-Carry (USD)	2.000

The bond price in six months is USD 92.214 and the investor receives a coupon of USD 2.000 just before the end of the six months. The initial bond price is USD 95.199. The gain (USD) is therefore:

$$92.214+2.000-95.199=-0.985$$

The P&L decomposition splits this into:

(a) A carry roll-down of USD 2.427,

(b) The impact of a term structure change of −USD 3.782, and

(c) A spread change of the USD 0.370.

The sum of (a), (b), and (c) is equal to the gain on the bond: $-0.985 = 2.427 - 3.782 + 0.370$

12

Applying Duration, Convexity, and DV01

Learning Objectives

After completing this reading, you should be able to:

- Describe a one-factor interest rate model and identify common examples of interest rate factors.
- Define and calculate the DV01 of a fixed-income security given a change in rates and the resulting change in price.
- Calculate the face amount of bonds required to hedge an interest rate-sensitive position given the DV01 of each.
- Define, calculate, and interpret the effective duration of a fixed-income security given a change in rates and the resulting change in price.
- Compare and contrast DV01 and effective duration as measures of price sensitivity.
- Define, calculate, and interpret the convexity of a fixed-income security given a change in rates and the resulting change in price.
- Calculate the DV01, duration, and convexity of a portfolio of fixed-income securities.
- Explain the hedging of a position based on effective duration and convexity.
- Construct a barbell portfolio to match the cost and duration of a given bullet investment and explain the advantages and disadvantages of bullet and barbell portfolios.

This chapter discusses several important concepts in the analysis of fixed-income portfolios: DV01, duration, and convexity. These are one-factor risk metrics (i.e., they are based on the assumption that interest rate term structure movements are driven by a single factor). In the next chapter, we will examine metrics that consider several such factors.

This chapter complements the chapter on interest rates in Financial Markets and Products. Whereas that chapter uses a yield-based definition of duration and convexity, this chapter focuses on effective duration and effective convexity. The yield-based measures consider what happens to a bond price when there is a small change to its yield. Effective duration and effective convexity consider what happens when all spot rates change by the same amount. We will refer to the latter as a *parallel shift* in the term structure.

Note that the measures considered in this chapter only quantify the impact of a parallel shift in the interest rate term structure. DV01 and duration describe what happens when there is a small parallel shift in the term structure. Convexity extends the duration analysis to larger parallel shifts. Hedging strategies based on these measures work well for parallel term structure shifts. However, they are liable to be less effective for non-parallel shifts.

12.1 THE ONE FACTOR ASSUMPTION

When rates are driven by a single factor, the movement of one interest rate over a short period of time can be used to determine the movement of all interest rates during that period. The simplest one-factor model assumes that all interest rates move by the same amount. Under this simple model, the shape of the term structure never changes. If the three-year spot rate increases by five basis points, all other spot rates increase by five basis points. If the three-year spot rate decreases by one basis point, all other spot rates decrease by one basis point. Figure 12.1 illustrates this model, which underlies the DV01, duration, and convexity measures considered in this chapter.

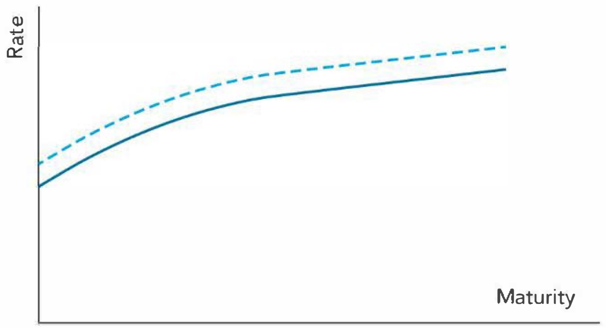


Figure 12.1 The solid line shows the initial term structure. The dashed line shows term structure after a parallel shift in rates.

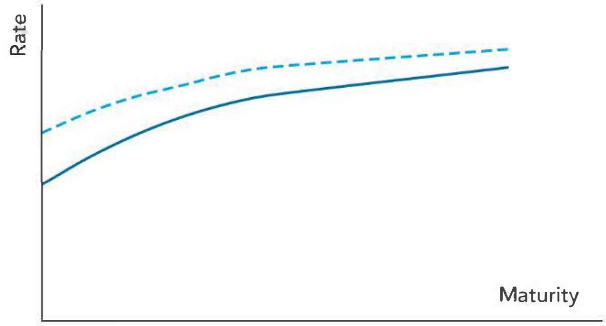


Figure 12.2 Solid line shows initial term structure. Dashed line shows term structure after a non-parallel shift in rates.

One-factor term structure shifts do not need to be parallel. In fact, some proposed one-factor models lead to long-maturity rates moving by less than short-maturity rates.[1] For example, a particular model might predict that when the one-year rate increases by ten basis points during a short period, the three-year rate increases by seven basis points and the ten-year rate increases by only four basis points. This type of model is illustrated in Figure 12.2. Note that we are still dealing with a one-factor model because the movement in all points over a short period can be calculated when the movement in one rate is known.

The term structure shape can change completely in a one-factor model. Suppose the short-maturity rates in Figure 12.2 continue to increase (and thus longer rates will increase by less). Eventually, the short-maturity rates will be greater than the long-maturity rates and the term structure will become downward-sloping (rather than upward-sloping). Similarly, the term structure will steepen if short-maturity rates decline sharply.

[1] The first one-factor equilibrium model of the term structure was proposed by O.A. Vasicek, "An equilibrium characterization of the term structure," *Journal of Financial Economics*, 5 (1977): 177–188. In this model, the change in the very short-maturity interest rate has a random normally distributed component and a component where the rate is pulled to a long-run average level. The latter is referred to as mean reversion and leads to the long-maturity rates moving by less than short-maturity rates. A no-arbitrage version of Vasicek's model that exactly fits the current term structure was proposed by J. Hull and A. White, "Pricing interest rate derivative securities," *Review of Financial Studies*, 3, 4 (1990): 573–592. Compared with other models that have been developed, the Vasicek and Hull-White models have the advantage that they give rise to many analytic results for bond prices and bond option prices.

12.2 DV01

DV01 describes the impact of a one-basis-point change in interest rates on the value of a portfolio. It is[2]

$$DV01 = -\frac{\Delta P}{\Delta r}$$

where Δr is the size of a small parallel shift in the interest rate term structure (measured in basis points) and ΔP is the resultant change in the value (P) of the position being considered.[3] For a long position in a bond, DV01 is positive because a bond price is negatively correlated to interest rate changes. The bond price increases (decreases) when rates decrease (increase).

For example, suppose a portfolio consists of a three-year Treasury bond with a face value of USD 1 million paying a 10% per annum coupon semi-annually. Suppose further that spot rates are as shown in Table 12.1. The value of the bond, assuming no spread, is

$$\frac{50{,}000}{1.035} + \frac{50{,}000}{1.0375^2} + \frac{50{,}000}{1.04^3} + \frac{50{,}000}{1.0415^4} + \frac{50{,}000}{1.0425^5} + \frac{1{,}050{,}000}{1.043^6} = 1{,}037{,}922.03$$

If rates increase by one basis point (i.e., by 0.01%), the six-month rate becomes 7.01%, the one-year rate becomes 7.51%, and so on. The value of the bond is then

$$\frac{50{,}000}{1.03505} + \frac{50{,}000}{1.03755^2} + \frac{50{,}000}{1.04005^3} + \frac{50{,}000}{1.04155^4} + \frac{50{,}000}{1.04255^5} + \frac{1{,}050{,}000}{1.04305^6} = 1{,}037{,}656.36$$

A one-basis-point increase in all rates causes the bond price to decline by USD 265.67 (= 1,037,922.03 − 1,037,656.36). An estimate of DV01 is therefore 265.67.

Now suppose that rates per year decrease by one basis point (i.e., the six-month rate becomes 6.99%, the one-year rate becomes 7.49%, and so on). The value of the bond becomes

$$\frac{50{,}000}{1.03495} + \frac{50{,}000}{1.03745^2} + \frac{50{,}000}{1.03995^3} + \frac{50{,}000}{1.04145^4} + \frac{50{,}000}{1.04245^5} + \frac{1{,}050{,}000}{1.04295^6} = 1{,}038{,}187.79$$

A one-basis point decrease in rates causes the bond price to increase by USD 265.76 (from USD 1,037,922.03 to USD 1,038,187.79). This provides an estimate of DV01 equal to 265.76. The DV01 estimate from considering a decrease in rates

Table 12.1 Spot Rates for Valuing Bond (Semi-Annually Compounded)

Maturity (Years)	Rate (%)
0.5	7.0
1.0	7.5
1.5	8.0
2.0	8.3
2.5	8.5
3.0	8.6

is almost exactly the same as that from considering an increase in rates. Averaging the two estimates we obtain

$$DV01 = 0.5 \times (265.67 + 265.76) = 265.72$$

The price of the bond is approximately linear in the change considered when the change is small. Instead of considering a one-basis-point change in rates, we could consider a five-basis point change and divide the change in the bond price by five. Table 12.2 summarizes the results from doing this. Averaging the two DV01 estimates, we again get 265.72.

The two estimates of DV01 in Table 12.2 differ slightly because the bond's price is not exactly a linear function of interest rates. We will return to this point later when we consider convexity.

In this definition of DV01, we have assumed all spot rates increase by one basis point. We get a slightly different definition of DV01 if we assume that the bond yield increases by one basis point. In our example, the bond yield is given by solving

$$\frac{50{,}000}{1+y/2} + \frac{50{,}000}{(1+y/2)^2} + \frac{50{,}000}{(1+y/2)^3} + \frac{50{,}000}{(1+y/2)^4} + \frac{50{,}000}{(1+y/2)^5} + \frac{1{,}050{,}000}{(1+y/2)^6} = 1{,}037{,}922.03$$

It is approximately 8.5404%. We now calculate a DV01 by increasing this yield by one basis point and decreasing it by one basis point. We find that:

- Increasing the yield by one basis point to 8.5504% decreases the bond price by USD 265.92, and
- Decreasing the yield by one basis point to 8.5304% increases the bond price by USD 266.00.

By averaging the two estimates, we get a DV01 estimate of 265.96.

Note that the DV01 per one-basis-point change in the bond yield (265.96) is only slightly different from the DV01 per one-basis-point change in all spot rates (265.72).

[2] The relationship between a bond price and rates is non-linear and so to be mathematically precise we should use calculus notation and define

$$DV01 = -\frac{dP}{dy}$$

[3] DV01 is also referred to as *dollar duration* when Δr is measured as a decimal rather than in basis points.

Table 12.2 Effect of Rate Changes on the Price of USD 1 Million Bond When Rates are Initially as in Table 12.1

Change	New Bond Price (USD)	Increase in Bond Price (USD)	DV01 Estimate
Five-basis-point increase in all rates	1,036,594.52	−1327.51	265.50
Five-basis-point decrease in all rates	1,039,251.68	+1329.65	265.93

Yet another possible definition of DV01 is obtained by increasing forward rates by one basis point. The result is again very similar to (but not quite the same as) increasing all spot rates by one basis point. Some analysts distinguish between these definitions as follows.

- *Yield-based DV01*: The change in price from a one-basis-point increase in a bond's yield.
- *DVDZ* or *DPDZ*: The change in price from a one-basis-point increase in all spot (i.e., zero) rates.
- *DVDF* or *DPDF*: The change in price from a one-basis-point increase in forward rates.

In this chapter we will assume that DV01 is calculated from a one-basis-point change in all spot rates (i.e., a one-basis-point parallel shift in the term structure of spot rates).

Hedging

DV01 can be calculated for any position whose value depends on interest rates. For example, suppose that the DV01 for a bank's position (in USD per basis point) is −463. This means the position gains value when interest rates increase and loses value when interest rates decrease.[4] Specifically:

- If all rates increase by one basis point, the value of the bank's position increases by USD 463, and
- If all rates decrease by one basis point, the value of the bank's position decreases by USD 463.

The bank's position can be hedged with the 10% coupon bond we have been considering.[5] Recall that when the bond's face value is USD 1 million, its DV01 is 265.72. If we establish a position in the bond so that its DV01 is 463, the DV01 of −463 for the bank's existing position will be exactly hedged.

To increase the bond's DV01 to 463, we need to increase the USD value of the position to:

$$1{,}000{,}000 \times \frac{463}{265.72} = 1{,}742{,}436$$

Adding a position of this size in the 10% coupon bond to the bank's portfolio hedges against small parallel shifts in the term structure. If the term structure moves up (down), there will be a gain (loss) on its existing position that will be offset by a loss (gain) on the position in the 10% coupon bond.

12.3 EFFECTIVE DURATION

As we have seen, DV01 is the decrease (increase) in the price of a bond (or other instrument) arising from a one-basis-point increase (decrease) in rates. When the instrument is a bond, $\Delta P/\Delta r$ is negative so that DV01 is positive.

While DV01 measures an actual change in price, effective duration describes the *percentage* change in the price of a bond (or other instrument) due to a small change in all rates. Denoting the effective duration of an instrument by D and the change in rates by Δr, we have:[6]

$$D = -\frac{\Delta P/P}{\Delta r} = -\frac{\Delta P}{P\Delta r} \tag{12.1}$$

This can be rewritten as:

$$\Delta P = -DP\Delta r \tag{12.2}$$

When the change in rates is measured in basis points, effective duration is DV01 divided by the price of the bond.

In the case of the bond considered in Section 12.2, $P = 1{,}037{,}922.03$ and $\Delta P/\Delta r = -265.72$ so that:

$$D = \frac{265.72}{1{,}037{,}922.03} = 0.000256$$

This is the proportional change in the price of the instrument due to a one-basis-point change in all interest rates. In this case, the effective duration is 0.0256%.

It is common to report effective duration as the percentage change in the price of an instrument for a 100-basis-point change in all rates. All this means is the impact of a one-basis-point change is multiplied by 100. In our example, effective duration would be reported as 2.56% per 100 basis points. Later in this

[4] This could be because it has entered into interest rate swaps where it is paying a fixed rate of interest and receiving a floating rate.

[5] This assumes that all interest rates, both those to which the bank is currently exposed, and those on which the value of the 10% coupon bond depend, move together in lock step.

[6] To be mathematically precise, we should use calculus notation and define

$$D = -\frac{1}{P}\frac{dP}{dy}$$

chapter, we will measure interest rates as decimals. One basis point is 0.0001 and therefore measuring interest rates in decimals means that we are considering effective duration per 10,000 basis points (and the duration in our example becomes 2.56).

It is important to emphasize that these are simply scaling issues. Whether we measure effective duration per basis point, per 100 basis points, or per 10,000 basis points, we are still considering the impact of a small change in interest rates on the value of an instrument. We do not calculate effective duration by observing the effect of a 100 or 10,000-basis-point change in rates because these are large changes. Instead, we calculate the effect of a small change and scale it up. For example, if we consider a five-basis-point change, as in Table 12.2, we would multiply the result by 20 to get a 100-basis-point quote, and by 2,000 to get a quote where rates are measured as decimals.

Note that we will measure duration with interest rates expressed in decimal terms in this chapter. The duration is therefore per 10,000 basis points and the duration for the example we have just considered is 2.56.

Callable and Puttable Bonds

A callable bond is a bond where the issuing company has the right to buy back the bond at a pre-determined price at certain times in the future.[7] A company will tend to do this when interest rates have declined so that it can re-finance at a lower rate.

Consider a five-year bond that can be called after three years (but not at any other time). There is a temptation to ignore the call feature when calculating effective duration and regard the bond as a non-callable five-year bond. This is an incorrect assumption because the call feature reduces duration.

Another approach would be to assume that the probability of the bond being called remains constant. For example, suppose that when the bond is valued there is a 40% chance that the bond will be called. When calculating the effective duration, we could calculate:

D_{called}: the effective duration of the three-year bond,

$D_{notcalled}$: the effective duration of the five-year bond

We could then set the effective duration of the callable bond as:

$$0.4D_{called} + 0.6D_{notcalled}$$

Although this approach is better than ignoring the call feature, it would not be correct either. When interest rates increase, the probability of the bond being called is reduced. A correct approach is therefore as follows.[8]

1. Value the bond today.
2. Value the bond if all interest rates increase by one basis point. (This calculation incorporates the effect of the one-basis-point increase on the probability of the bond being called.)
3. Calculate effective duration from the percentage change in the price.

A puttable bond is a bond where the holder has the right to demand early repayment. A puttable bond should be treated like a callable bond when calculating effective duration. In this case, the probability of the put option being exercised increases as interest rates increase.

Effective Duration versus DV01

In choosing between effective duration and DV01, an analyst must decide whether to consider the impact of rate changes on the value of a position in dollars or as a percentage. In the first case, DV01 is appropriate; in the second case, effective duration is appropriate. DV01 increases as the size of a position increases, while effective duration does not. (If a position is doubled in size, DV01 doubles while effective duration remains the same.)

A bond investor is usually interested in returns, which are typically measured in percentage terms. This usually means effective duration is the better measure.

In other situations, DV01 may be more appropriate. For example, consider a bank that has just entered an interest rate swap. We can calculate how the value of the swap would change for a one-basis-point change in rates. However, effective duration is not a meaningful measure in this situation because the value of the swap is zero (or close to zero). Therefore, DV01 is likely to be the most appropriate measure for swaps. (This is also the case for interest rate futures.)

12.4 CONVEXITY

Effective convexity measures the sensitivity of the duration measure to changes in interest rates. The effective convexity (C) of a position worth P can be estimated as:[9]

$$C = \frac{1}{P}\left[\frac{P^+ + P^- - 2P}{(\Delta r)^2}\right] \tag{12.3}$$

[7] Governments and other entities (e.g., special purpose vehicles) can also issue callable bonds.

[8] Binomial trees can be used to value bonds with embedded options. The use of binomial trees is explained in Chapter 14 of this book.

[9] Using calculus terminology, convexity is $\frac{1}{P}\frac{d^2P}{dy^2}$. The expression $\frac{P^+ + P^- - 2P}{(\Delta y)^2}$ is a numerical approximation to $\frac{d^2P}{dy^2}$. To see this, note that it can be written $\frac{(P^+ - P)/\Delta y - (P - P^-)/\Delta y}{\Delta y}$.

where P^+ is the value of the position when all rates increase by Δr and P^- is the value of a position when all rates decrease by Δr. As in the definition of duration, we measure Δr as a decimal.

Consider again the bond in Section 12.2. This is a three-year bond paying a coupon of 10% per year semi-annually. The interest rates in the market are those shown in Table 12.1, and the bond price (P) is USD 1,037,922.03. When Δr equals 0.0005 (or five basis points), we know from the calculation in Section 12.2 (see Table 12.2) that:

$$P^+ = 1{,}036{,}594.52$$

$$P^- = 1{,}039{,}251.68$$

The estimate of convexity given by Equation (12.3) is therefore:

$$C = \frac{1}{1{,}037{,}922.03} \times \left[\frac{1{,}036{,}594.52 + 1{,}039{,}251.68 - 2 \times 1{,}037{,}922.03}{0.0005^2}\right]$$
$$= 8.246$$

The Impact of Parallel Shifts

Recall that effective duration provides the impact of a small parallel shift in the term structure. Consider again the bond in Section 12.2 valued at USD 1,037,922.03. If rates increase by 20 basis points, the term structure is as shown in Table 12.3:

Table 12.3 Rates in Table 12.1 After a Parallel Shift of 20 Basis Points

Maturity (Years)	Rate (%)
0.5	7.2
1.0	7.7
1.5	8.2
2.0	8.5
2.5	8.7
3.0	8.8

The value of the bond (USD) is

$$\frac{50{,}000}{1.036} + \frac{50{,}000}{1.0385^2} + \frac{50{,}000}{1.041^3} + \frac{50{,}000}{1.0425^4} + \frac{50{,}000}{1.0435^5} + \frac{1{,}050{,}000}{1.044^6}$$
$$= 1{,}032{,}624.79$$

The bond's price decreases by USD 5,297.24 (= 1,037,922.03 − 1,032,624.79). From Equation (12.2), the duration relation predicts a price change of:

$$-DP\Delta r = -2.56 \times 1{,}037{,}922.03 \times 0.002 = -5{,}314.16$$

This is reasonably accurate, but it can be made more accurate using the convexity (C). The estimate of the price change using both duration and convexity is[10]

$$-DP\Delta r + \frac{1}{2}CP(\Delta r)^2 \tag{12.4}$$

In this case, $C = 8.246$ and so we get a revised estimate of the price change as:

$$-5314.16 + \frac{1}{2} \times 8.246 \times 1{,}037{,}922.03 \times 0.002^2 = -5{,}297.04$$

This is very close to the price decrease of 5,297.24 that we calculated by revaluing the bond.

The estimate's accuracy declines as the size of the parallel shift considered increases. However, the duration + convexity approximation can give a reasonable answer even for quite large changes. For example, consider a 200-basis-point increase in all rates for the bond in our example. The bond's price declines to USD 986,448.71 (a decrease in value of USD 51,473.32). Using duration alone indicates a price change of:

$$-2.56 \times 1{,}037{,}922.03 \times 0.02 = -53{,}141.61$$

But using the duration + convexity result in Equation (12.4) gives

$$-53{,}141.61 + \frac{1}{2} \times 8.246 \times 1{,}037{,}922.03 \times 0.02^2 = -51{,}429.04$$

which is reasonably accurate.

The effective duration approximation for a bond is illustrated in Figure 12.3. Note that there is a non-linear relationship between $\Delta P/P$ and Δr.[11] Effective duration approximates this as a linear relation because it is the gradient of the curve relating $\Delta P/P$ and Δr at the origin. The linear approximation provided by duration becomes less accurate as the (positive or negative) magnitude of Δr increases. The duration plus convexity approximation fits a quadratic function and captures some of the curvature. As a result, it provides a better approximation.

One important point illustrated by Figure 12.3 is that the actual percentage change in a bond price is always greater than that predicted by duration (i.e., duration overestimates the magnitude of price decreases and underestimates the magnitude of

[10] This is based on the Taylor formula approximation:

$$\Delta P = \frac{dP}{dr}\Delta r + \frac{1}{2}\frac{d^2P}{dr^2}(\Delta r)^2 \text{ and the definitions } D = -\frac{1}{P}\frac{dP}{dr} \text{ and } C = \frac{1}{P}\frac{d^2P}{dr^2}$$

[11] For clarity, the non-linearity typically encountered has been exaggerated in Figure 12.3.

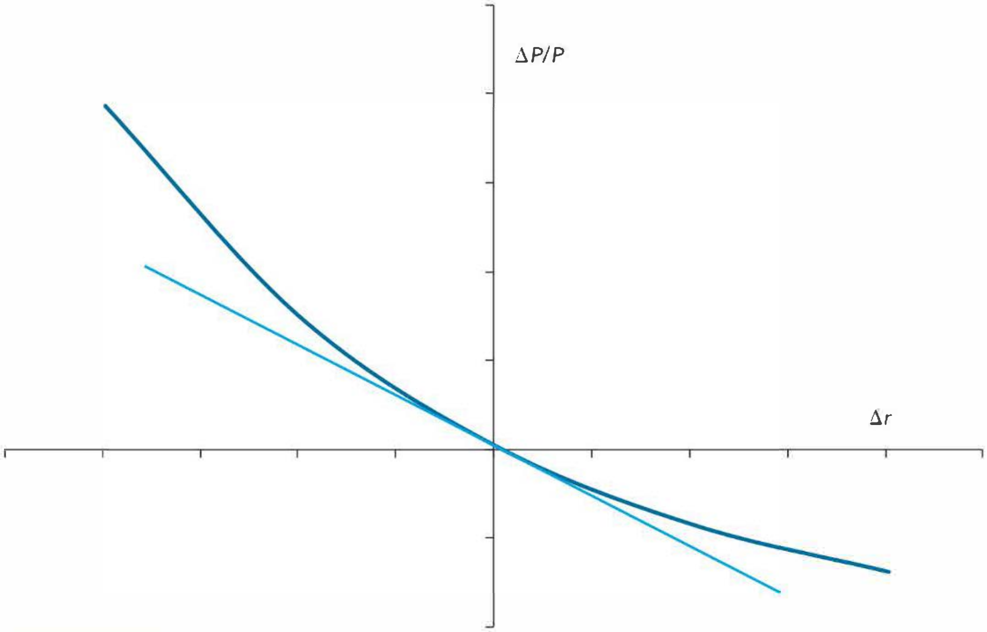


Figure 12.3 **Impact of change Δr in all rates.**

price increases). Meanwhile, the convexity measure provides an estimate of the difference between:

- The bond price change predicted by effective duration, and
- The actual bond price change.

When convexity is positive, the impact of a parallel shift in rates on the value of a bond portfolio is better than duration alone would predict. When convexity is negative, the impact of a parallel shift in rates on the value of the portfolio is worse than duration alone would predict.

As convexity increases, the curvature of the relationship between bond price changes (measured in percentage terms) and rate changes increases. It can be seen from Figure 12.3 that this improves the return on a bond when there is a parallel shift in rates. We will return to this point when we consider bullet and barbell investments in Section 12.7.

Hedging

Hedging based on effective duration is like hedging based on DV01 (considered in Section 12.2). Suppose that the effective duration of an investment is D_V and the value of the investment is V. Suppose further that the duration of a bond is D_P and the value of the bond is P. Because:

$$\Delta V = -VD_V\Delta r$$

and

$$\Delta P = -PD_P\Delta r$$

where Δr is the size of a small parallel shift in the term structure, we are hedged against small parallel shifts if:

$$-VD_V - PD_P = 0$$

The position required in the bond is therefore:

$$P = -\frac{VD_V}{D_P}$$

For example, suppose we have a position worth USD 2 million with an effective duration of 4.0 and a bond with an effective duration of 5.0. To hedge the position, the value of the position taken in the bond (in USD million) should be

$$-\frac{2 \times 4}{5} = -1.6$$

Therefore, bonds worth USD 1.6 million should be shorted. It is then the case that:

$$\Delta V = -2 \times 4 \times \Delta r = -8\Delta r$$

$$\Delta P = 1.6 \times 5 \times \Delta r = +8\Delta r$$

For a more sophisticated hedging approach, we can try to make both effective duration and effective convexity zero. For this we need two bonds. Suppose P_1, D_1, and C_1 are the value, duration, and convexity of the first bond (respectively), while P_2, D_2, and C_2 are the value, duration, and convexity of the second bond (also respectively). If V, D_V, and C_V are the value, duration, and convexity of the position that is to be hedged, Equation (12.4) gives

$$\Delta V = -VD_V\Delta r + \frac{1}{2}VC_V(\Delta r)^2$$

$$\Delta P_1 = -P_1D_1\Delta r + \frac{1}{2}P_1C_1(\Delta r)^2$$

$$\Delta P_2 = -P_2D_2\Delta r + \frac{1}{2}P_2C_2(\Delta r)^2$$

We can make both duration and convexity zero by choosing P_1 and P_2 so that:

$$\begin{aligned} -VD_V - P_1D_1 - P_2D_2 &= 0 \\ VC_V + P_1C_1 + P_2C_2 &= 0 \end{aligned} \qquad (12.5)$$

To continue our previous example (where values are measured in USD million), assume $V = 2$ and $D_V = 4$. We also suppose $C_V = 12$, $D_1 = 5$, $C_1 = 10$, $D_2 = 3$, and $C_2 = 8$. Equations (12.5) become

$$8 + 5P_1 + 3P_2 = 0$$

$$24 + 10P_1 + 8P_2 = 0$$

and the solution is $P_1 = 0.8$ and $P_2 = -4$. If a position of USD 0.8 million in the first bond is combined with a short position of USD 4 million in the second bond, there is no duration or convexity exposure. This means that the position is hedged against relatively large parallel shifts in the term structure. However, it will still have exposure to non-parallel shifts.

12.5 YIELD-BASED DURATION AND CONVEXITY

In the interest rates chapter of Financial Markets and Products, we introduced the yield-based duration and convexity measures. For completeness, we now review the properties of these measures.

Consider a bond with price P and yield y. Because it leads to a simpler duration formula, we first measure the yield with continuous compounding. If cash flows $c_1, c_2, \ldots$, and c_n are received by the bond holder at times $t_1, t_2, \ldots$, and t_n (respectively), the relationship between P and y is

$$P = \sum_{i=1}^{n} c_i e^{-yt_i}$$

It is then true that:[12]

$$\Delta P = -\sum_{i=1}^{n} c_i t_i e^{-yt_i} \Delta y$$

We define the yield-based duration as the proportional change in the bond price for a small change in the yield. The yield-based duration is

$$D = \frac{1}{P}\sum_{i=1}^{n} t_i c_i e^{-yt_i} = \sum_{i=1}^{n} t_i \frac{c_i e^{-yt_i}}{P} \tag{12.6}$$

The expression:

$$\frac{c_i e^{-yt_i}}{P}$$

denotes the proportion of the bond's value received at time t_i. This analysis therefore gives us another interpretation of duration. It can be calculated by taking an average of the times when cash flows are received weighted by the proportion of the bond's value received at each time.[13] Duration is therefore a measure of how long an investor has to wait to receive cash flows. (This explains why the word "duration" was chosen to describe the sensitivity of proportional changes to yield).

[12] This is obtained by differentiating the expression for P:

$$\frac{dP}{dy} = -\sum_{i=1}^{n} c_i t_i e^{-yt_i}$$

[13] Values are here calculated by using the yield rather than spot rates for discounting.

If (as is usual) the yield on the bond is measured with semi-annual compounding rather than continuous compounding, the expression for duration in Equation (12.6) must be divided by $1 + y/2$ so that it becomes[14]

$$D = \frac{1}{P(1 + y/2)}\sum_{i=1}^{n} t_i c_i e^{-yt_i} = \frac{1}{1 + y/2}\sum_{i=1}^{n} t_i \frac{c_i e^{-yt_i}}{P}$$

This is referred to as *modified duration*, whereas the duration in Equation (12.6) is sometimes referred to as Macaulay's duration (because it was suggested by Frederick Macaulay in 1938).

Yield-based convexity when yields are expressed with continuous compounding is

$$C = \frac{1}{P}\sum_{i=1}^{n} t_i^2 c_i e^{-yt_i} = \sum_{i=1}^{n} t_i^2 \frac{c_i e^{-yt_i}}{P}$$

It is a weighted average of the squared time to maturity. When yields are expressed with semi-annual compounding, these expressions must be divided by $(1 + y/2)^2$ and the result is referred to as modified convexity.

Earlier in this chapter we showed that DV01 calculated from a one-basis-point change in a yield is only slightly different from DV01 calculated from a one-basis-point change in all spot rates. Similarly, modified duration is only slightly different from effective duration and modified convexity is only slightly different from effective convexity. When a yield-based measure (DV01, duration, or convexity) for one bond is compared to that of another for building a hedge, it is implicitly assumed that all yields move in lock step.

12.6 PORTFOLIO CALCULATIONS

We now consider the situation where a portfolio consists of a number of bonds (or other instruments) and we wish to calculate an interest-rate-sensitivity measure for the portfolio using the same measure for the instruments within the portfolio.

Consider first DV01. The DV01 for a portfolio is simply the sum of the DV01s of the components of the portfolio. If a portfolio consists of three positions that have DV01s (in thousands of USD) of 30, 40, and 50 (respectively), the DV01 of the portfolio is 120 (= 30 + 40 + 50).

Now consider duration. The duration of a portfolio is the average of the durations for the individual instruments within the portfolio weighted by the value of each instrument. For example, consider a portfolio consisting of three bonds worth

[14] More generally, if the yield is measured with a compounding frequency of m times per year, we must divide by $1 + y/m$.

Table 12.4 Effective Durations and Convexities of Three Bonds

Bond	Value	Effective Duration	Effective Convexity
5-year, 2% coupon	91.0174	4.6764	24.8208
10-year, 4% coupon	100.0000	8.1758	78.8981
20-year, 6% coupon	127.3555	12.6235	212.4604

(in USD millions) 10, 15, and 25. Suppose that the effective durations of the bonds are 6.0, 8.0, and 11.0 (respectively). The effective duration of the portfolio is

$$\frac{10}{10+15+25} \times 6 + \frac{15}{10+15+25} \times 8 + \frac{25}{10+15+25} \times 11$$
$$= 0.2 \times 6 + 0.3 \times 8 + 0.5 \times 11 = 9.1$$

This is the duration that could be used to calculate the impact a small parallel shift in the interest rate term structure on the value of a portfolio. A similar calculation can be used for yield-based duration measures, which measure the change in a portfolio's value given a small parallel shift in the yields on all bonds within the portfolio. For example, consider a portfolio that consists of two bonds: one with a yield of 3% and one with a yield of 4%. The yield-based duration measure could be used to show the effect of both the first yield increasing to 3.01% and the second yield increasing to 4.01%.

Convexity for a portfolio is handled like duration. It is the average of the convexities of the instruments within the portfolio weighted by the value of each instrument.

12.7 BARBELL VERSUS BULLET

Consider three bonds:

1. A 5-year bond with a 2% coupon,
2. A 10-year bond with a 4% coupon, and
3. A 20-year bond with a 6% coupon.

We assume that the term structure of interest rates is flat at 4% (semi-annually compounded). The effective durations and convexities of the three bonds can be calculated as indicated earlier in this chapter. They are shown in Table 12.4. (Interest rates are measured in decimals for the calculation of effective duration and effective convexity.)

If an investor wants a portfolio with an effective duration of 8.1758, he or she can buy the ten-year, 4% coupon bonds. This is referred to as a *bullet investment* because only one bond is involved. An alternative is to construct a portfolio from the other two bonds with an effective duration equal to 8.1758. This is known as a *barbell investment*. Suppose that a proportion β of the portfolio is invested in the five-year bond, and a proportion $(1-\beta)$ is invested in the 20-year bond. The duration will be

$$4.6764\beta + 12.6235(1-\beta)$$

This equals 8.1758 when β is 0.5597 (which we will round to 0.56).

There are therefore two ways a portfolio with a duration of 8.1758 can be created

1. Invest all funds in the ten-year, 4% coupon bond, or
2. Invest 56% of funds in the five-year, 2% coupon bond and 44% of the funds in the 20-year, 6% coupon bond.

The portfolios have the same duration but different convexities. The first has a convexity of 78.8981. The second alternative has a convexity of:

$$24.8208 \times 0.56 + 212.4604 \times 0.44 = 107.382$$

As illustrated by Figure 12.3, a positive convexity improves the bond holder's position when there is a parallel shift in interest rates. As convexity increases, the improvement increases.

While both strategies provide a yield of 4% and a duration of 8.1758, the barbell strategy always produces a better result when there is a parallel shift in the yield curve. The barbell strategy therefore appears to dominate the bullet strategy.

We might deduce from this that there is an opportunity for arbitrageurs:

- Invest a certain USD amount in the barbell, and
- Short the same USD amount of the bullet

This would be profitable if shifts in the term structure were always parallel. However, this is not the case and non-parallel shifts do occur. In fact, the bullet investment performs better than the barbell investment for many non-parallel shifts in the term structure.[15]

It is also worth noting that we have assumed that the yield curve is initially flat at 4% (so that the yield on all instruments is 4%).

[15] For example, if the term structure becomes upward-sloping after one year so that the yields on the 5-year, 10-year, and 20-year bonds are 3%, 4%, and 5%, respectively, the value of the bullet position is unchanged, but the value of the barbell position has declined by about 2%.

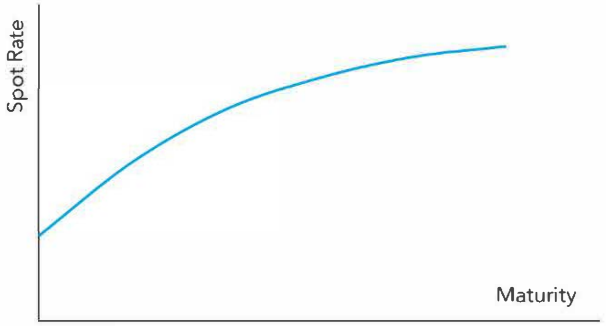


Figure 12.4 **Typical upward sloping term structures.**

This means that if the term structure remains unchanged, all three bonds will provide a 4% return and thus the returns from the barbell and bullet investments will be the same. However, if the yield curve instead has the commonly observed upward-sloping convex shape (indicated in Figure 12.4), it can be shown that the yield on the bullet investment is greater than the yield on the barbell investment.[16] If yields stay the same, the bullet investment will outperform the barbell investment by the yield differential.

When researchers construct models of how the interest rate term structure moves, their objective is usually to produce what is termed a no-arbitrage model. This is a model where there are no arbitrage opportunities open to investors. Suppose a simple model is proposed where the term structure is always flat. Based on our arguments in this section, we know that a simple model where the term structure is always flat is not a no-arbitrage model.

[16] Question 12.18 at the end of this chapter provides an illustration of this.

SUMMARY

DV01, effective duration, and effective convexity provide information about what will happen if there is a parallel shift in the interest rate term structure. DV01 is the decrease (increase) in portfolio value for a one-basis-point increase (decrease) in all rates. Effective duration similarly measures the sensitivity of the percentage change in the value of a portfolio to a small change in all rates. Meanwhile, effective convexity measures the sensitivity of effective duration to a small change in all rates.

In the case of bonds, we can define DV01, duration, and convexity in terms of small changes in yields rather than small changes in all rates. This leads to some analytic results and explains the name "duration."

The DV01 of a portfolio is the sum of the DV01s of the instruments in the portfolio. The duration (convexity) of a portfolio is the average of the durations (convexities) of the instruments in the portfolio weighted by the value of each instrument.

Two portfolios with the same value and the same duration can have different convexities. The portfolio with the higher convexity will always perform better when there is a parallel shift in interest rates. However, it may not do so for nonparallel shifts. For some commonly observed initial term structures, the low convexity portfolio has a higher yield and will provide a better return than the high convexity portfolio when the yields do not change.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

QUESTIONS

Short Concept Questions

12.1 What are three ways that DV01 can be defined?

12.2 How is effective duration defined?

12.3 What is a one-factor interest rate model?

12.4 What is a callable bond? Does it become more or less likely to be called when interest rates increase?

12.5 When the size of a position is doubled, what happens to (a) DV01, (b) duration, and (c) convexity?

12.6 Suppose that the duration of a bond position is zero and convexity is positive. Does the value of the bond position increase or decrease when there is a small parallel shift in rates?

12.7 What protection is provided by (a) DV01 hedging, (b) effective duration hedging, and (c) effective duration plus convexity hedging?

12.8 How can yield-based duration be calculated for a bond without revaluation?

12.9 Why is yield-based convexity likely to be greater than yield-based duration for a ten-year bond (assume that rates are expressed with continuous compounding)?

12.10 What is the difference between a bullet and a barbell investment?

Practice Questions

12.11 Consider a zero-coupon bond with a face value of USD 100 and a maturity of ten years. What is the DV01 and the effective duration when the ten-year rate is 4% with semi-annual compounding? (Consider one-basis-point changes and measure rates as decimals when calculating duration.)

12.12 What is the effective convexity of the bond in Question 12.11? (Consider one-basis-point changes and measure rates as decimals.)

12.13 What is the difference between the effective duration, modified duration, and the Macaulay's duration for the bond in Question 12.11?

12.14 An investor has a bond position worth USD 20,000 with a duration of 7.0. How can the position be hedged with a bond that has a duration of 10.0?

12.15 Suppose that the bond position in Question 12.14 has a convexity of 33. Two bonds are available for hedging. One has a duration of 10.0 and a convexity of 80. The other has a duration of 6.0 and a convexity of 25. How can a duration plus convexity hedge be set up?

12.16 What is the effective duration and convexity of a three-year Treasury bond with a face value of 1 million and a coupon of 4% when the term structure is flat at 5%? Express interest rates in decimals and consider five-basis-point changes.

12.17 Estimate the effect of all rates in Question 12.16 increasing by 0.25% using (a) duration and (b) duration plus convexity.

12.18 Suppose that the five-, ten-, and 30-year rates are 4%, 5%, and 6% with semi-annual compounding. Calculate the duration and convexity of zero-coupon bonds with five-, ten-, and 30-years to maturity. What position in five- and 30-year bonds would have a duration equal to that of the ten-year bond? Compare the convexities of (a) the positions in the ten-year bond and (b) the position in the five- and 30-year bonds? Which of these positions will give the better return if (a) rates remain the same and (b) there are parallel shifts in the term structure?

12.19 A portfolio consists of three instruments:

1. An instrument worth 10.0 with a duration of 3.0,
2. An instrument worth 6.0 with a duration of 5.0, and
3. An instrument worth 4.0 with a duration of 7.0.

What is the duration of the portfolio?

12.20 A position is worth USD 1.5 million. A two-basis-point increase in all rates causes the value to decline by USD 1199.85 and a two-basis-point decrease in all rates cause the value to increase by USD 1200.15. Estimate the effective duration and effective convexity.

ANSWERS

Short Concept Questions

12.1 DV01 can be defined as the reduction in price for (a) a one-basis-point increase in all spot rates or (b) a one-basis-point increase in yield or (c) a one-basis-point increase in forward rates.

12.2 Effective duration for an instrument with price P is

$$-\frac{\Delta P}{P\Delta r}$$

where Δr is the size of a small increase in all rates and ΔP is the resultant change in the price.

12.3 A one-factor interest rate model is a model where rate changes are driven by a single factor. If we know how one rate has moved during a short period of time we can deduce how all rates have moved during that period.

12.4 A callable bond is a bond where the issuer can choose to repay the amount borrowed early. The issuer is less likely to do this when rates rise because the cost of refinancing the bond becomes greater.

12.5 DV01 doubles. The duration and convexity remain the same because they reflect percentage price changes.

12.6 It increases. Duration alone predicts no change, but a positive convexity leads to an increase in value when there is a parallel shift.

12.7 DV01 and effective duration hedging provide protection against small parallel shifts in the term structure. Effective duration plus convexity provides protection against larger parallel shifts.

12.8 Macaulay's yield-based duration is the weighted average of the times when cash flows are received with the weight applied to time t being proportional to the present value of the cash flow at time t. This is a correct yield-based duration measure when rates are expressed with continuous compounding. It must be divided by $(1 + y/2)$ where y is the yield when rates are expressed with semi-annual compounding.

12.9 Yield-based convexity is calculated by squaring each cash flow's time to maturity and then taking a weighted average with weights proportional to the present values of the cash flows. Yield-based duration is a weighted average of the time to maturity of cash flows with the same weights. For a ten-year bond, the former is clearly greater than the later because, for nearly all the cash flows, the square of the time to payment is greater than the time to payment.

12.10 A bullet is an investment in a single bond. A barbell is an investment in a portfolio consisting of a short-maturity and a long-maturity bond.

Solved Problems

12.11 The value of the bond is

$$\frac{100}{1.02^{20}} = 67.297133$$

When the ten-year rate increases to 4.01%, the value decreases by 0.065944 to 67.231190. When the ten-year rate decreases to 3.99%, the value increases by 0.066012 to 67.363145. The DV01 can be estimated as the average of 0.065944 and 0.066012, or 0.065978. The effective duration is

$$\frac{0.065978}{67.297133 \times 0.0001} = 9.804$$

12.12 The effective convexity is

$$\frac{67.231190 + 67.363145 - 2 \times 67.297133}{67.297133 \times 0.0001^2} = 100.9$$

Note that more decimal places than those indicated were kept to provide this estimate of convexity.

12.13 Macaulay's duration is ten years because the bond only provides a cash flow at the ten-year point. The effective duration is 9.804 years. The modified duration is the Macaulay's duration divided by $1 + 0.04/2$. This is 9.804. Because this is a zero-coupon bond, the effective duration and the modified duration are the same. (Making a small change to the yield has the same effect as making a small parallel shift in all rates.)

12.14 The position (USD) required in bond is

$$-\frac{20{,}000 \times 7}{10} = -14{,}000$$

12.15 To make both duration and convexity equal to zero, we must solve

$$10P_1 + 6P_2 + 20{,}000 \times 7 = 0$$
$$80P_1 + 25P_2 + 20{,}000 \times 33 = 0$$

This gives $P_1 = -2{,}000$ and $P_2 = -20{,}000$, indicating that a short position of USD 2000 in the first bond and a short position of USD 20,000 in the second bond is required.

12.16 The value of the bond is 97.245937. When there is five-basis-point increase in all rates so that the term structure is flat at 5.05%, the value falls by 0.135287 to 97.110650. When there is a five-basis-point decrease in all rates so that the term structure is flat at 4.95%, the value rises by 0.135514 to 97.381452. The duration is

$$\frac{0.5 \times (0.135287 + 0.135514)}{97.245937 \times 0.0005} = 2.784703$$

The convexity is

$$\frac{97.110650 + 97.381452 - 2 \times 97.245937}{97.245937 \times 0.0005^2} = 9.35$$

Note that even more decimal places than those indicated is necessary to provide this estimate of convexity.

12.17 The predicted change using duration is

$$-97.245937 \times 2.784703 \times 0.0025 = -0.677003$$

Using duration plus convexity we get the estimated change as

$$-97.245937 \times 2.784703 \times 0.0025 + \frac{1}{2} \times 97.245937$$
$$\times 9.349608 \times 0.0025^2 = -0.674161$$

The actual change is -0.674170, very close to this.

12.18 The duration and convexities calculated by making one-basis-point changes are

Bond Maturity	Duration	Convexity
5	4.902	26.423
10	9.756	99.941
30	29.126	862.476

We can construct a bond with a duration of 9.756 by investing β in the five-year bond and $1 - \beta$ in the 30-year bond where:

$$4.902\beta + 29.126(1 - \beta) = 9.756$$

β is 0.7996, which we round to 0.8. We therefore invest 80% in the five-year bond and 20% in the 30-year bond. The ten-year bond investment (a bullet) has a convexity of 99.941 whereas the portfolio of five- and 30-year bonds (a barbell) has a convexity of about:

$$0.8 \times 26.423 + 0.2 \times 862.472 = 193.6$$

If rates remain the same the bullet will provide a yield of 5%, whereas the barbell will provide a weighted average yield of $0.8 \times 4 + 0.2 \times 6$ or 4.4%. The bullet will perform better. When there are parallel shifts to the term structure, this effect is mitigated somewhat by the barbell's higher convexity, which leads to an immediate improvement in the value of the barbell position. However, the bullet will perform better for some non-parallel shifts.

12.19 The duration is

$$3 \times \frac{10}{20} + 5 \times \frac{6}{20} + 7 \times \frac{4}{20} = 4.4$$

12.20 Effective duration is

$$\frac{0.5 \times (1{,}199.85 + 1{,}200.15)}{1{,}500{,}000 \times 0.0002} = 4$$

Also:

$$P^+ + P^- - 2P = (P^+ - P) - (P - P^-)$$
$$= 1200.15 - 1195.85 = 3$$

so that effective convexity is

$$\frac{0.3}{1{,}500{,}000 \times 0.0002^2} = 5$$

13

Modeling Non-Parallel Term Structure Shifts and Hedging

Learning Objectives

After completing this reading, you should be able to:

- Describe principal components analysis and identify the factors that are the most important drivers of term structure movements.
- Describe key rate shift analysis and define key rate 01 (KR01).
- Calculate the KR01s of a portfolio given a set of key rates.
- Calculate the positions in hedging instruments necessary to hedge the key rate risks of a portfolio.
- Apply key rate analysis and principal components analysis to estimating portfolio volatility.
- Describe an interest rate bucketing approach, define forward bucket 01, and compare forward bucket 01s to KR01s.
- Calculate the corresponding duration measure given a KR01 or forward bucket 01.

The risk measures introduced in Chapter 12 consider only parallel shifts in the interest rate term structure. In practice, there are many different types of non-parallel shifts. Sometimes short-term rates move down while long-term rates move up, or vice versa. Occasionally, short- and long-term interest rates move in one direction, while medium-term rates move in the other direction. Hedging based on DV01, duration, and convexity (as described in Chapter 12) leaves an investor exposed to these types of movements.

In this chapter, we extend the metrics introduced in Chapter 12 to provide models that can hedge a wide range of different term structure movements. We also explain how these models can be used to calculate value-at-risk (VaR) for a portfolio of interest-rate-sensitive instruments. As we will explain, bank regulators require banks to use one of these models to calculate market risk capital and initial margin.

13.1 PRINCIPAL COMPONENTS ANALYSIS

A statistical technique known as *principal components analysis* can be used to understand term structure movements in historical data. This technique looks at the daily movements in rates of various maturities and identifies certain factors. These factors are term structure movements with the property that:

- The daily term structure movements observed are a linear combination of the factors (e.g., an observed movement might consist of five units of the first factor, two units of the second factor, one unit of the third factor, and so on);
- The factors are uncorrelated; and
- The first two or three factors account for most of the observed daily movements.

The material in Chapter 12 assumed just one factor (a parallel shift in the term structure). A principal components analysis finds multiple factors and estimates their relative importance in describing movements in the term structure.

Principal components analysis is best illustrated with an example. The Federal Reserve Board provides daily data on Treasury rates with maturities of 1, 2, 3, 5, 7, 10, 20, and 30 years.[1] We can use this data from the period between January 2008 and December 2019 to carry out a principal components analysis on the daily changes in these eight rates. The number of factors equals the number of rates.[2] What are known as *factor loadings* are shown in Table 13.1. These are the amount by which each of the rates moves when there is one unit of the factor.[3]

The factors are listed in order of importance (Factor 1 is the most important factor, Factor 2 is next most important, and so on). When there is +1 unit of Factor 1, the one-year rate changes by −0.134 basis points, the two-year rate changes by −0.266 basis points, and so on. When there is +1 unit of Factor 2 the one-, two-, three-, and five-year rates move down while the seven-, ten-, 20-, and 30-year rates move up. When there is +1 unit of Factor 3, the first two and last two rates move up while the intermediate four rates move down.

Note that if we change all the signs of all the factor loadings for a particular factor, it does not change the analysis. In Table 13.1, all the factor loadings for Factor 1 are negative. This means that +1 unit of Factor 1 will cause all rates to decrease and −1 unit of Factor 1 will cause all rates to increase. Now suppose that we change all the signs so that the factor loadings for the rates with maturities 1, 2, 3, 5, 7, 10, 20, and 30 years are 0.134, 0.266, 0.331, 0.410, 0.432, 0.411, 0.383, and 0.364 (respectively). It is then the case that +1 unit of the new Factor 1 has the same effect as −1 unit of the old Factor 1 (and vice versa). Because +1 unit and −1 unit of a factor are equally likely, the model is unchanged.

Because there are eight rates and eight factors, the change on any particular day can be calculated as a linear combination of the factors. The change in the *j*th rate has the form:

$$\sum_{i=1}^{8} a_i f_{ij}$$

where f_{ij} is the factor loading for the *i*th factor and the *j*th rate and a_i $(1 \leq i \leq 8)$ is the number of units of the *i*th factor in the daily change being considered.

The a_i are referred to as *factor* scores. They are different for each of the daily changes in the historical sample.

[1] See www.federalreserve.gov/data. The rates are estimates of the yields on bonds with these maturities.

[2] Any changes in *n* rates can be expressed as a linear combination of *n* factors by solving *n* simulataneous linear equations. This explains why an analysis involving *n* rates gives rise to *n* factors.

[3] For more on principal components analysis, see J. Hull, "Risk Management and Financial Institutions," fifth edition, Wiley 2018. Software for carrying out principal component analysis calculations is available at www-2.rotman.utoronto.ca/~hull/riskman.

Table 13.1 Factor Loadings for the Treasury Rates Estimated from Data Between January 2008 and December 2019

Rate Maturity	Factor 1	2	3	4	5	6	7	8
1 year	−0.134	−0.385	0.786	0.465	0.002	0.002	−0.015	0.006
2 year	−0.266	−0.493	0.067	−0.599	0.568	−0.009	0.025	0.001
3 year	−0.331	−0.420	−0.129	−0.219	−0.736	0.294	0.135	−0.058
5 year	−0.410	−0.195	−0.314	0.237	−0.060	−0.589	−0.529	0.094
7 year	−0.432	0.015	−0.286	0.394	0.228	−0.074	0.709	−0.128
10 year	−0.411	0.196	−0.101	0.195	0.221	0.692	−0.346	0.311
20 year	−0.383	0.399	0.230	−0.164	−0.024	−0.005	−0.188	−0.761
30 year	−0.364	0.444	0.344	−0.317	−0.174	−0.287	0.209	0.544

The importance of a factor is measured by the standard deviation of its factor scores. For the data we are considering, the importance of the *i*th factor is the standard deviation of a_i across all the approximately 3,000 daily changes in the term structure during the 2008–2019 period. The standard deviation of the factor scores are shown in Table 13.2.

Additionally, the variances of the factor scores add up to the total variance of all rate movements. For our data, the total variance is

$$13.58^2 + 4.66^2 + \cdots + 0.66^2 = 216.44$$

The first factor accounts for 85.1% ($= 13.58^2/216.44$) of the variance. The first two factors account for:

$$\frac{13.58^2 + 4.66^2}{216.44} = 95.16\%$$

of the variance. Similarly, the first three factors account for 97.64% of the variance.

This shows that most of the uncertainty in interest rate movements can be accounted for by using the first three factors. These first three factors are plotted in Figure 13.1. The factors can be seen to have the following properties.

Table 13.2 Standard Deviation of Factor Scores for Each Factor

Factors 1	2	3	4	5	6	7	8
13.58	4.66	2.32	1.54	1.05	0.82	0.74	0.66

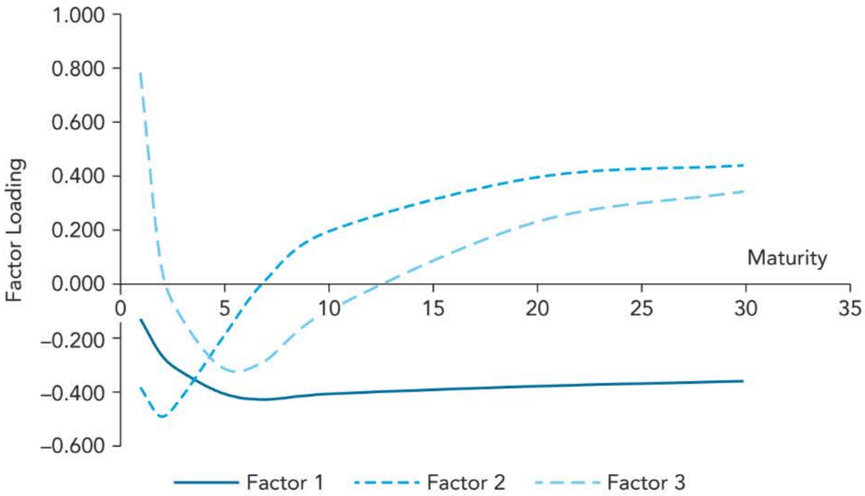


Figure 13.1 The three most important factors driving Treasury rates.

- Factor 1 is a shift in the term structure where all rates move in the same direction by approximately (but not exactly) the same amount.[4]
- Factor 2 is a shift where short-term rates move in one direction and long-term rates move in the other direction. It corresponds to steepening or flattening of the term structure.
- Factor 3 is a bowing of the term structure (i.e., where relatively short-term and relatively long-term rates move in one direction while intermediate rates move in the other direction).

[4] It might seem surprising that the one-year rate moves by less than the other rates in the first two factors. This is because one-year rates stayed very low during most of the period considered and tended to move by less than other rates.

This analysis shows that the one-factor (parallel shift) model considered in Chapter 12 provides an imperfect hedging tool since it only approximately reflects movements in the first factor (which accounts for about 85% of rate changes). In this chapter, we consider how the metrics introduced in Chapter 12 can be extended to accommodate a multi-factor model.

13.2 PARTIAL 01S

In Chapter 12, we explained that DV01 can be defined as the impact of a one-basis-point shift in all spot rates on the value of a portfolio. Suppose we consider three spot rates: the two-year rate, the five-year rate, and the ten-year rate. Figures 13.2, 13.3, and 13.4 show how shifts in each of those spot rates can be defined. The shifts in Figures 13.2 to 13.4 are sometimes referred to as *key rate shifts*.

The combined effect of the three key rate shifts in Figures 13.2 to 13.4 is a one-basis-point shift in all rates. These figures therefore provide a way of splitting the DV01 measure used in Chapter 12 into three other measures.

The impact of shifts, such as the ones shown in Figures 13.2 through 13.4, are sometimes referred to as *partial 01s* or *key rate 01s* (KR01s).

Define

$KR01_1$: The reduction in a portfolio's value from a one-basis-point increase in the two-year spot rate, as in Figure 13.2

$KR01_2$: The reduction in a portfolio's value from a one-basis-point increase in the five-year spot rate, as in Figure 13.3

$KR01_3$: The reduction in a portfolio's value from a one-basis-point increase in the ten-year spot rate, as in Figure 13.4

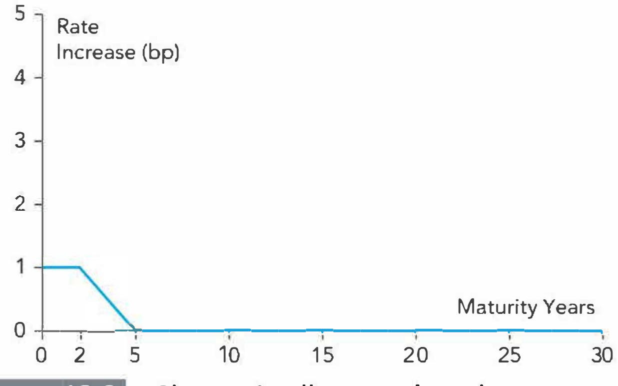


Figure 13.2 Change in all rates when the two-year rate is increased by one basis point.

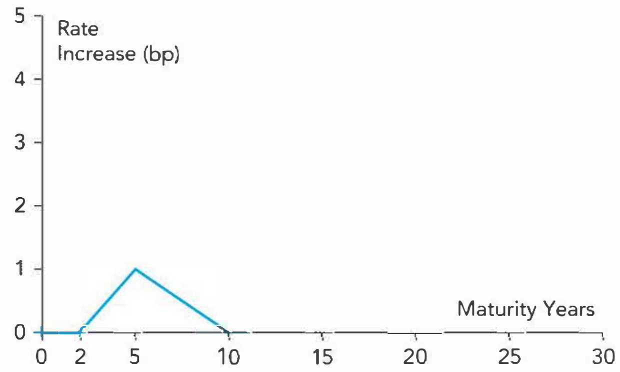


Figure 13.3 Change in all rates when the five-year rate is increased by one basis point.

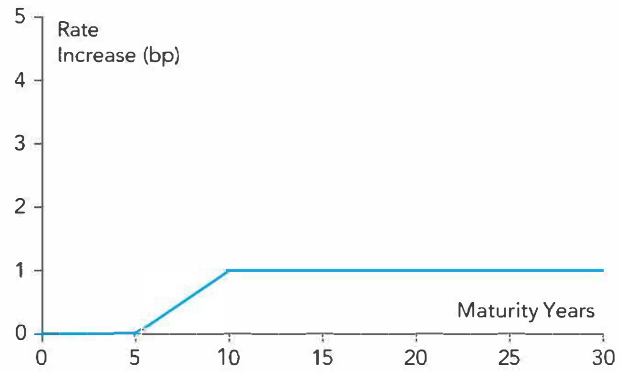


Figure 13.4 Change in all rates when the ten-year rate is increased by one basis point.

It follows that:

$$DV01 = KR01_1 + KR01_2 + KR01_3$$

In Chapter 12, we discussed how an investor can use DV01 to hedge against parallel shifts in the interest rate term structure. If an investor follows a similar procedure to hedge against $KR01_1$, $KR01_2$, and $KR01_3$, the investor is hedged against a wider range of term structure movements. Specifically, the investor is hedged against small rate movements of the type shown in Figures 13.2 through 13.4 as well as any combination of these three rate movements.

KR01s (either for a portfolio or for instruments used to hedge a portfolio) can usually be calculated in a fairly direct way from their sensitivities to spot rates. To take a simple example, suppose a portfolio consists of a USD 1 million investment in each of a one-year, three-year, five-year, nine-year, and 15-year zero coupon bond. Suppose further that the term structure is flat at 3% (semi-annually compounded). The decrease in the portfolio's value for a one-basis-point increase in the relevant spot rates is shown in Table 13.3.

Table 13.3 Decrease in Value of Portfolio for a One-Basis-Point Increase in Spot Rates

Spot Rate Maturity (Years)	1	3	5	9	15
Portfolio Value Decrease	95.62	270.26	424.35	677.93	944.74

The exposures in Table 13.3 can be converted to exposures to the shifts in Figures 13.2, 13.3, and 13.4. Table 13.4 shows the impact of the shifts in Figures 13.2, 13.3, and 13.4 on the five spot rates in Table 13.3. For example, the shift in Figure 13.2 involves a one-basis-point shift in the one-year rate and a 0.6667-basis-point shift in the three-year rate, with none of the other rates being affected.

The KR01s for the portfolio with the exposures in Table 13.3 can therefore be calculated as indicated in Table 13.5.

To illustrate how a hedge position can be obtained, we will use the data in Table 13.6. This shows KR01s for a portfolio and

Table 13.4 Changes in Spot Rate for Changes in Figures 13.2, 13.3, and 13.4

	Spot Rate Maturity (Yrs)				
Shift	**1**	**3**	**5**	**9**	**15**
Figure 13.2	1	0.6667	0	0	0
Figure 13.3	0	0.3333	1	0.2	0
Figure 13.3	0	0	0	0.8	1

Table 13.5 KR01s for Portfolio

Partial 01	Calculation	Result
$KR01_1$	$95.62 + (0.6667 \times 270.26)$	275.8
$KR01_2$	$(0.3333 \times 270.26) + 424.35 + (0.2 \times 677.93)$	650.0
$KR01_3$	$(0.8 \times 677.93) + 944.74$	1,487.1

Table 13.6 Data for Hedging Using KR01s

		Hedging Instruments		
	Portfolio	**1**	**2**	**3**
$KR01_1$	126	20	3	3
$KR01_2$	238	2	22	4
$KR01_3$	385	1	4	25

three different hedging instruments. The positions in the three hedging instruments necessary to reduce the KR01s to zero can be calculated by solving three simultaneous equations. If x_1, x_2, and x_3 are the positions in the three hedging instruments, the equations are

$$126 + 20x_1 + 3x_2 + 3x_3 = 0$$
$$238 + 2x_1 + 22x_2 + 4x_3 = 0$$
$$385 + x_1 + 4x_2 + 25x_3 = 0 \tag{13.1}$$

The solution to these equations is $x_1 = -3$, $x_2 = -8$, and $x_3 = -14$. The portfolio can therefore be hedged with short positions of 3, 8, and 14 in the three hedging instruments.

Bank Regulation

Bank regulators require banks to analyze the risks in their portfolios by considering ten different KR01s. Specifically, they require banks to calculate the impact of one-basis-point shifts in the three-month, six-month, one-year, two-year, three-year, five-year, ten-year, 15-year, 20-year, and 30-year spot rates. The shifts are calculated in the way indicated in Figures 13.2 to 13.4 so that the sum of the KR01s always equals the DV01. Banks do not attempt to make all the KR01s zero, but they are required to calculate risk measures (VaR or expected shortfall) using their KR01 exposures in conjunction with the standard deviations of, and correlations between, the ten rates specified by regulators. The risk measures are used to (a) calculate capital for market risk under new rules being introduced by the Basel Committee, and (b) calculate initial margin for derivative transactions not routed through a central counterparty.[5] The formula for the standard deviation (σ_P) of the change in value of the portfolio in one day is

$$\sigma_P = \sqrt{\sum_{i=1}^{10} \sum_{j=1}^{10} \rho_{ij}\sigma_i\sigma_j \times KR01_i \times KR01_j} \tag{13.2}$$

where σ_i is the standard deviation of the daily movement in rate i (measured in basis points) and ρ_{ij} is the correlation between the daily movements in rate i and rate j. If the change in the value of the portfolio can be assumed to be normal, this leads directly to VaR and expected shortfall measures.

A Generalization

We can generalize the ideas introduced in this section so that any set of term structure movements are considered. When there are n term structure movements, n simultaneous equations must be solved to perfectly hedge against the movements.

[5] See Bank for International Settlements, "Minimum Capital Requirements for Market Risk," January 2016 and "ISDA SIMM: From Principles to Model Specification," ISDA, March 3, 2016.

In practice, hedging is often less than perfect. Suppose that the exposure to the ith term structure movement is w_i. Equation (13.2) for the standard deviation of the value of the portfolio becomes

$$\sigma_P = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_i \sigma_j \rho_{ij} w_i w_j} \tag{13.3}$$

The estimate of σ_P improves as the set of term structure movements considered provides a more complete description of the term structure movements that occur in practice.

13.3 PRINCIPAL COMPONENTS ANALYSIS REVISITED

We now return to the principal components analysis in Section 13.1. We showed that the first three factors account for about 97.6% of the variance in Treasury rate movements over a ten-year period. One approach to hedging is to work with these three factors.

To implement the hedge, it is necessary to produce a table like Table 13.6 with the three term structure movements being those in Figure 13.1. As in Equation (13.1), three simultaneous equations must be solved to determine the hedge. The hedge provides protection not only against the movements in Figure 13.1, but also against movements that are any combination of those in Figure 13.1 (e.g., a movement that consists of 3 units of the first factor, −2 units of the second factor, and +1 unit of the third factor). Assuming the ten-year history is representative of future term structure movements, the portfolio is hedged against 97.6% of the variance in term structure movements.

The shifts given by a principal components analysis in Figure 13.1 involve more rates than the shifts in Figures 13.2 to 13.4. Furthermore, they are not linear combinations of those shifts. The protection provided by the hedges in Section 13.2 is therefore likely to be less comprehensive than the protection provided by hedges based on the factors coming out of a principal components analysis.

It is particularly easy to use the factors in Figure 13.1 to calculate VaR and expected shortfall because these factors are uncorrelated. Define σ_i as the standard deviation of the factor score for the ith factor (as given by Table 13.2), and f_i as the change in the value of the portfolio when there is a movement in the term structure corresponding to one unit of the ith factor. From equation (13.3), the standard deviation of the daily change in the value of the portfolio based on the first three factors is

$$\sigma_P = \sqrt{\sigma_1^2 f_1^2 + \sigma_2^2 f_2^2 + \sigma_3^2 f_3^2}$$

Suppose an analysis shows the following.

- The portfolio value changes by +20 when the term structure has the changes indicated by the first factor in Table 13.1.
- The portfolio value changes by +35 when the term structure has the changes indicated by the second factor in Table 13.1.
- The portfolio value changes by −10 when the term structure has the changes indicated by the third factor in Table 13.1.

Using the results in Table 13.2, the standard deviation of the daily changes in the value of the portfolio is

$$\sqrt{13.58^2 \times 20^2 + 4.66^2 \times 35^2 + 2.32^2 \times 10^2} = 317.52$$

If we assume a normal distribution, the ten-day 99% VaR is estimated as:

$$\sqrt{10} \times N^{-1}(0.99) \times 317.52 = 2{,}336$$

The ten-day 99% expected shortfall is

$$\sqrt{10} \times 317.52 \frac{e^{-[N^{-1}(0.99)]^2/2}}{\sqrt{2\pi} \times 0.01} = 2{,}676$$

13.4 THE USE OF PAR YIELDS

In Section 13.2, we defined the key rate shifts in Figures 13.2 to 13.4 as changes to spot rates. An alternative is to define key rate shifts in terms of par yields. This has the advantage that we can immediately calculate the position necessary to hedge a portfolio once we have calculated the exposure of the portfolio to the key rate shifts.[6]

We can consider the three key shifts in Figures 13.2 to 13.4, but now assume that they describe changes in par yields rather than spot rates (i.e., the vertical axis is now "change in par yield"). The three changes add up to a yield-based DV01.[7] Suppose that for a portfolio we find the following.

- A one-basis-point increase in the two-year par yield causes the portfolio value to increase by 20.
- A one-basis-point increase in the five-year par yield causes the portfolio value to increase by 30.
- A one-basis-point increase in the ten-year par yield causes the portfolio value to increase by 35.

The portfolio can be hedged with a position in a two-year par yield bond with a yield-based DV01 of 20, a position in a

[6] The use of par yields was suggested by T. Ho, "Key Rate Duration: A Measure of Interest Rate Risk," *Journal of Fixed Income*, Sept 1992: 29-44.

[7] As indicated in Chapter 12, making a one-basis-point change in all spot rates is much the same as making a one-basis-point change in all yields.

five-year par yield bond with a yield-based DV01 of 30, and a position in a ten-year par yield bond with a yield based DV01 of 35. The DV01s are actually KR01s with respect to par yields rather than spot rates.

The term structure of spot rates is typically calculated from the market prices of actively traded instruments; in the fixed income market, these are typically par yield (or close-to-par yield) bonds. This means that the yields on par yield bonds define the term structure of spot rates and the term structure of spot rates can change only if one of the yields changes.

Calculating a portfolio's exposure to par yields is therefore not too difficult in many situations. The impact of a change in a particular par yield on the spot rate term structure can be calculated by recomputing spot rates after making the change.[8] The new spot rates can then be used to calculate a new portfolio value. The difference between this and the current portfolio value indicates the exposure to a change in that par yield.

In the swap market, a trader typically measures his or her exposure to swap rates. This is analogous to a trader in the fixed income market determining exposure to par yields. Calculations used to determine exposures are then like those we have described for the fixed-income market. As explained in the swaps chapter of Financial Markets and Products, it is now common to value swaps using overnight index swap (OIS) discounting so that two term structures are involved. However, this makes the hedging calculations more complicated.

13.5 BUCKETING APPROACH

A variation on the key rate shifts approach is to divide the interest rate term into segments referred to as *buckets,* and then calculate the dollar impact of changing all the spot rates in a bucket by one basis point on the value of a portfolio.

Figures 13.5, 13.6, and 13.7 show three bucket shifts that could be used instead of the three key rate shifts in Figures 13.2, 13.3, and 13.4. Define B_1, B_2, and B_3 as the impact on a portfolio value for the shifts in Figures 13.5, 13.6, and 13.7 (respectively). As in the case of the key rate shifts, the sum of the one-basis-point shifts in these figures is the DV01 of the portfolio:

$$\text{DV01} = B_1 + B_2 + B_3$$

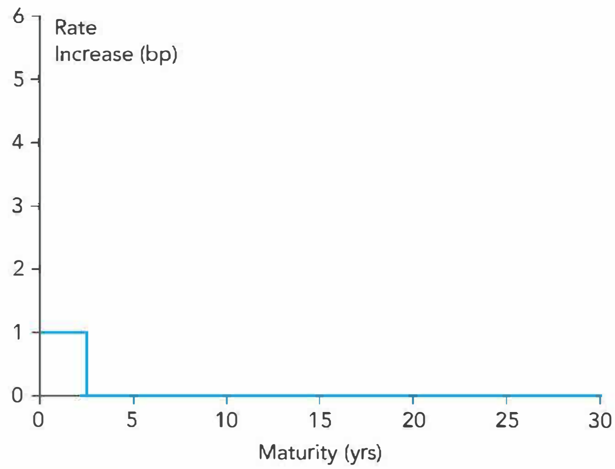


Figure 13.5 One-basis-point change in rates for 0–2 year bucket.

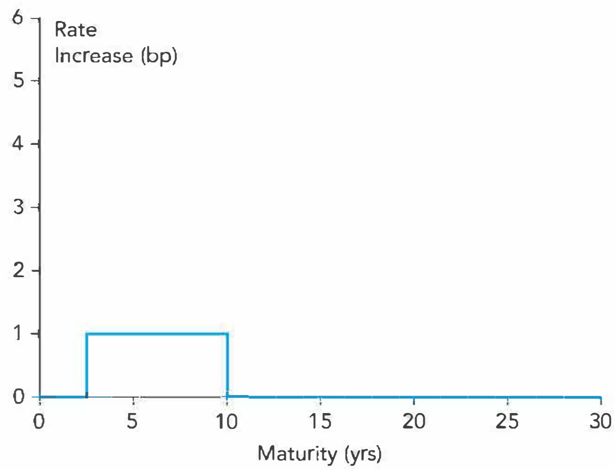


Figure 13.6 One-basis-point change in rates for 2–10 year bucket.

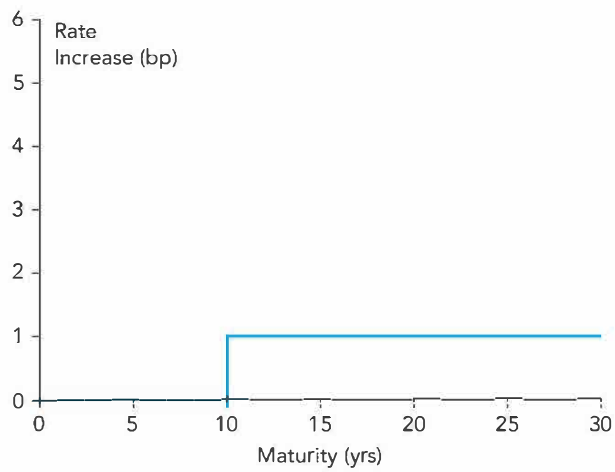


Figure 13.7 One-basis-point change in rates for 10–30 year bucket.

[8] The bootstrap method for calculating spot rates from instruments observed in the market was described in the "Interest Rates" chapter of Financial Markets and Products.

This bucketing approach is often used by banks in asset-liability management. However, more than three buckets are usually used. A bank is reasonably well-hedged if, for each bucket, the decline in the value of assets and the decline in value of liabilities are approximately the same for a one-basis-point increase in rates.[9]

13.6 FORWARD BUCKET SHIFTS

The bucketing procedure we have just described for spot rates can be used for forward rates as well. Suppose we bucket rates as described in the previous section so that there are three buckets: 0–2 years, 2–5 years, and 5–10 years. Assume forward rates are calculated for six-month periods. Applying a one-basis-point shift to forward rates in the first bucket would involve increasing the forward rates for the six-month periods beginning in 0, 6, 12, and 18 months by one basis point. Similarly, applying a one-basis-point shift to forward rates in the second bucket would involve increasing the forward rates for the six-month periods beginning in 24, 30, 36, 42, 48, and 54 months by one basis point. Applying a one-basis-point shift to forward rates in the third bucket would involve increasing the six-month forward rates starting in 60, 66, 72, . . . , months by one basis point. The decrease in portfolio value for a one-basis-point increase in all forward rates in a bucket is referred to as a *forward bucket 01*. The sum of these forward bucket 01s equals the DV01 calculated by changing forward rates.

Hedging using forward-rate shifts is similar to the other hedging procedures we have outlined in this chapter. We must calculate the sensitivity of the portfolio and the sensitivity of the hedging instruments to the forward rate shifts, and then solve simultaneous equations analogous to the equations in Equation (13.1).

Spot rates can be calculated from forward rates. Suppose all rates are six-month rates expressed with semi-annual compounding. Define f_t as the forward rate for the period between t and $t + 0.5$ years. The N-year spot rate (R) is then given by:

$$(1 + 0.5R)^{2N} = (1 + 0.5f_0)(1 + 0.5f_{0.5})(1 + 0.5f_1)\ldots(1 + 0.5f_{N-0.5}) \quad (13.4)$$

This formula can be used to calculate the impact of one-basis-point changes in forward rates on spot rates and therefore on the value of a portfolio or hedging instrument.

Consider a simple portfolio consisting of a three-year bond with a face value of USD 100 and a coupon of 6% per year. Assume the term structure is flat at 4% with semi-annual compounding.

[9] A major category of interest-rate sensitive assets is fixed-rate loans. A major category of interest-rate sensitive liabilities is fixed-rate term deposits.

The value of the bond is

$$\frac{3}{1.02} + \frac{3}{1.02^2} + \frac{3}{1.02^3} + \frac{3}{1.02^4} + \frac{3}{1.02^5} + \frac{103}{1.02^6} = 105.6014$$

Suppose there are three buckets: 0–1 years, 1–2 years, and 2–3 years. When forward rates in the 0–1 year bucket are increased by one basis point, the value of the bond becomes

$$\frac{3}{1.02005} + \frac{3}{1.02005^2} + \frac{3}{1.02005^2 \times 1.02} + \frac{3}{1.02005^2 \times 1.02^2} + \frac{3}{1.02005^2 \times 1.02^3} + \frac{103}{1.02005^2 \times 1.02^4} = 105.5912$$

so that the forward bucket 01 is 0.0102 (= 105.6014 − 105.5912). When forward rates in the 1–2 year bucket increase by one basis point, the value of the bond becomes

$$\frac{3}{1.02} + \frac{3}{1.02^2} + \frac{3}{1.02^2 \times 1.02005} + \frac{3}{1.02^2 \times 1.02005^2} + \frac{3}{1.02^2 \times 1.02005^2 \times 1.02} + \frac{103}{1.02^2 \times 1.02005^2 \times 1.02^2} = 105.5918$$

so that the forward bucket 01 is 0.0096 (= 105.6014 − 105.5918). When forward rates in the 2–3 year bucket increase by one basis point, the value of the bond becomes

$$\frac{3}{1.02} + \frac{3}{1.02^2} + \frac{3}{1.02^3} + \frac{3}{1.02^4} + \frac{3}{1.02^4 \times 1.02005} + \frac{103}{1.02^4 \times 1.02005^2} = 105.5923$$

so that the forward bucket 01 is 0.0091 (= 105.6014 − 105.5923). These results are summarized in Table 13.7.

The total of the three forward bucket 01s is 0.0289. This is also the result of increasing all the forward rates by one basis point because:

$$\frac{3}{1.02005} + \frac{3}{1.02005^2} + \frac{3}{1.02005^3} + \frac{3}{1.02005^4} + \frac{3}{1.02005^5} + \frac{103}{1.02005^6} = 105.5725$$

which is 105.6014 − 0.0289.

In our example, the forward rates in the first bucket have a larger effect on price than the forward rates in the second bucket, which in turn have a larger effect on forward rates than the forward rates in the third bucket. This illustrates a general phenomenon: The early forward rates affect more cash flows than the later forward rates, and therefore have higher 01s.

As we saw in the swaps chapter of Financial Markets and Products, a swap is a portfolio of forward rate agreements. The value of each forward rate agreement depends on the current forward rate. It is therefore natural for swap traders to calculate

Table 13.7 Forward Bucket 01s for a Three-Year, 6% Coupon Bond with a Face Value of USD 100 When the Term Structure is Flat at 4% (Compounded Semi-Annually)

Bucket	Bond Price with One-Basis-Point Shift in the Bucket's Forward Rates	Forward Bucket 01
0–1 Year	105.5810	0.0102
1–2 Years	105.5821	0.0096
2–3 Years	105.5832	0.0091
Total		0.0289

their exposures to forward rates in the way we have described. Forward bucket 01s are also useful for contracts such as swaptions (i.e., options to enter a particular swap in the future).

13.7 DURATION MEASURES

So far, we have considered how a DV01 (defined in terms of spot rates, par yields, or forward rates) can be decomposed in several measures that add up to the DV01. This can also be done with duration. We know from Chapter 12 that when interest rates are measured as decimals in the calculation of duration:

$$\text{Duration} = \frac{10{,}000 \times \text{DV01}}{\text{Value of Portfolio}}$$

We can similarly convert any of the 01 measures presented in this chapter into a duration measure using:

$$\text{Duration Measure} = \frac{10{,}000 \times (01\,\text{Measure})}{\text{Value of Portfolio}}$$

Consider the measures produced in the example in Section 13.6. We can convert the forward bucket 01 measures in Table 13.7 to forward bucket duration measures as indicated in Table 13.8. For example, the 0–1 year forward bucket duration is

$$\frac{10{,}000 \times 0.0102}{105.6014} = 0.97$$

Table 13.8 shows how the total duration of 2.74 can be split into three components.

Table 13.8 Calculation of Forward Bucket Durations from Results in Table 13.6

Bucket	Forward Bucket 01	Forward Bucket Durations
0–1 Year	0.0102	0.97
1–2 Years	0.0096	0.91
2–3 Years	0.0091	0.86
Total	0.0289	2.74

A similar approach can be used to convert (a) the spot rate KR01s in Section 13.2, (b) the par yield KR01s in Section 13.4, and (c) the bucketed 01s in Section 13.5.

SUMMARY

The measures introduced in Chapter 12 describe the impact of a parallel shift in the interest rate term structure. In practice, perfectly parallel shifts are rare. A principal components analysis shows the term structure changes observed can be approximated by three factors:

- A factor where all rates move in the same direction, but not by exactly the same amount.
- A factor where the term structure steepens or flattens.
- A factor where there is a bowing of the term structure.

It therefore makes sense for analysts to consider more term structure changes than the simple parallel shift.

In Chapter 12, we mentioned there are three different ways of defining DV01. These involve changing (a) spot rates, (b) bond yields, and (c) forward rates. In this chapter, we have shown how each of these DV01s can be split into several component measures. We have also shown how the sensitivity of a portfolio to shifts in the term structure can be used to calculate the standard deviation of the daily change in the portfolio value, and therefore provide estimates of risk measures such as VaR and expected shortfall.

Finally, we have extended the DV01 analysis so that it applies to duration. As a result, any duration measure (whether calculated by considering spot rates, bond yields, or forward rates) can be split into several components, each corresponding to a simple shift in the term structure.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

QUESTIONS

Short Concept Questions

13.1 In a principal components analysis, what is (a) factor loading and (b) a factor score?

13.2 Explain two ways key rate shifts can be defined.

13.3 What is the difference between a KR01 and a DV01?

13.4 In what respect is it easier to calculate the standard deviation of the daily change in a portfolio's value by using the factors produced by a principal components analysis than by using KR01s?

13.5 What is the advantage of using par yield KR01s rather than spot rate KR01s for hedging?

13.6 What is the difference between the key rate 01 approach and the bucketing approach?

13.7 What is a forward bucket 01?

13.8 Explain why the forward bucket 01s corresponding to the first bucket of a term structure tend to be larger than those corresponding to the last bucket of a term structure.

13.9 What is the difference between a 01 measure and the corresponding duration measure?

13.10 How is the par yield KR01 measure for a portfolio converted into a duration measure?

Practice Questions

13.11 Suppose a portfolio has an exposure of +50 to a one-basis-point increase in the five-year Treasury rate in Table 13.1, an exposure of −100 to a one-basis-point increase in the ten-year Treasury rate in Table 13.1, and no other exposures. What is the portfolio's exposure to the first two factors in Table 13.1?

13.12 Using Table 13.2, calculate the standard deviation of the daily change in the portfolio in Question 13.11 based on its exposure to the first two factors.

13.13 What is the estimated 20-day, 95% VaR for the portfolio in Question 13.11?

13.14 Suppose that the 12-month and 30-month spot rates are chosen as key rates. Plot the key rate 01 shifts.

13.15 What are KR01s for an 18-month, zero-coupon bond with a face value of USD 1,000 when the key rates are as in Question 13.14? Assume the 18-month rate is 5% with semi-annual compounding.

13.16 Why do banks tend to use KR01s calculated from spot rates?

13.17 Suppose par yield KR01s are calculated using five- and ten-year shifts in par yields. A portfolio has an exposure of +20 to a one-basis-point change in the seven-year par yield. Use linear interpolation to determine its par yield KR01s.

13.18 Two hedging instruments are available with the exposures shown in the following table. What positions in the instruments should be taken to zero out the exposure of the portfolio in Question 13.17 to the five- and ten-year key rate shifts?

	Hedging Instrument 1	Hedging Instrument 2
KR01 (Five-Year Shift)	4	2
KR01 (Ten-Year Shift)	2	2

13.19 Calculate the forward bucket 01s for a two-year bond with a coupon of 8% and a face value of USD 10,000 when the there are two buckets: 0–1 year and 1–2 year. Assume that the term structure is flat at 4% (semi-annually compounded).

13.20 Convert the forward bucket 01s in Question 13.19 to durations.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

ANSWERS

Short Concept Questions

13.1 A factor loading shows the basis point movement in an interest rate when there is one unit of the factor. The factor score is the number of units of a factor in a daily term structure change.

13.2 Key rate shifts can be defined in terms of spot rate changes or in terms of par yield changes. They have the form indicated in Figures 13.2 to 13.4.

13.3 A DV01 is the result of shifting all rates by one basis point. A KR01 is a result of shifts in key rates such as those shown in Figure 13.2 to 13.4.

13.4 The factors produced by a principal components analysis are uncorrelated with each other, whereas the KR01 shifts are not uncorrelated.

13.5 When par yield KR01s are used, it is easy to identify how par yield bonds can be used as hedges.

13.6 The KR01 approach shifts a particular rate by one basis point (with other nearby rates moving by less than one basis point). (See Figures 13.2 to 13.4.) The bucketing approach shifts all rates in a certain range by one basis point. (See Figures 13.5 to 13.7.)

13.7 A forward bucket 01 is the decrease in the value of a portfolio when all the forward rates within a particular bucket (i.e., all the forward rates corresponding to a particular range of maturities) are increased by one basis point.

13.8 Changes to the forward rates in the first bucket affect more spot rates (and therefore more cash flows) than the forward rates in the last bucket.

13.9 The 01 measure shows the dollar change in the value of the portfolio for a small change in the term structure. The corresponding duration measure shows the percentage change in the value of the portfolio for a small change in the term structure. The 01 measure is a per basis point change. For the duration measure, interest rates are measured as decimals.

13.10 It is multiplied by 10,000 and divided by the value of the portfolio.

Solved Problems

13.11 The exposure to one unit of the first factor is

$$50 \times (-0.410) - 100 \times (-0.411) = 20.6$$

The exposure to one unit of the second factor is

$$50 \times (-0.195) - 100 \times 0.196 = -29.35$$

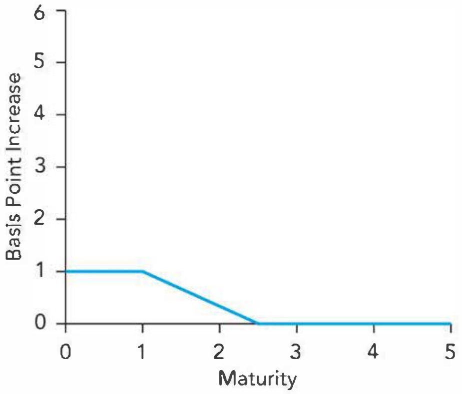
13.12 Using Table 13.2, the standard deviation is

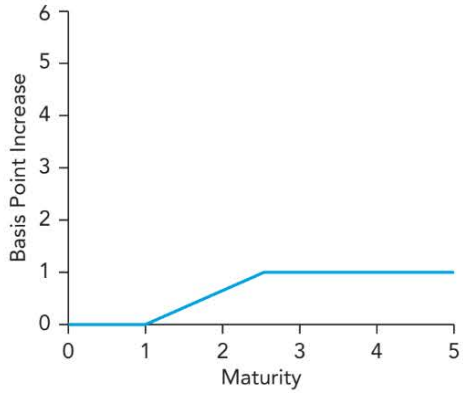
$$\sqrt{13.58^2 \times 20.6^2 + 4.66^2 \times 29.35^2} = 311.39$$

13.13 The VaR is

$$\sqrt{20} \times N^{-1}(0.95) \times 311.39 = 2{,}290.59$$

13.14 The Key Rate shifts are





13.15 The bond's price is 928.5994. When the 18-month rate increases by one basis point, the price decreases to 928.4635. The decrease in price is 0.1359. The KR01 for the 12-month rate shift is $0.1359 \times 0.667 = 0.0906$. The KR01 for the 30-month rate shift is $0.1359 \times 0.333 = 0.0453$.

13.16 They are used for (a) the new Basel rules for determining capital for market risk and (b) rules for determining initial margin for transactions not cleared through central counterparties (CCPs)s.

13.17 The KR01 for the five-year key rate is $0.6 \times 20 = 12$. The KR01 for the ten-year key rate is $0.4 \times 20 = 8$.

13.18 Suppose that x_1 and x_2 are the positions in the two hedging instruments. We require

$$4x_1 + 2x_2 + 12 = 0$$
$$2x_1 + 2x_2 + 8 = 0$$

The solution to these equations is $x_1 = -2$ and $x_2 = -2$. We need to take a short position of two in each hedging instrument.

13.19 The value of the bond is

$$\frac{400}{1.02} + \frac{400}{1.02^2} + \frac{400}{1.02^3} + \frac{10{,}400}{1.02^4} = 10{,}761.5457$$

When the forward rates in the first bucket increase by one basis point, the value of the bond becomes

$$\frac{400}{1.02005} + \frac{400}{1.02005^2} + \frac{400}{1.02005^2 \times 1.02} + \frac{10{,}400}{1.02005^2 \times 1.02^2} = 10{,}760.5100$$

This is a decrease of 1.0358. When the forward rates in the second bucket increase by one basis point, the value of the bond becomes

$$\frac{400}{1.02} + \frac{400}{1.02^2} + \frac{400}{1.02^2 \times 1.02005} + \frac{10{,}400}{1.02^2 \times 1.02005^2} = 10{,}760.5854$$

This is a decrease of 0.9604. The forward bucket 01s are therefore 1.0358 and 0.9604.

13.20 The duration measure for the first forward bucket is

$$\frac{10{,}000 \times 1.0358}{10{,}761.5457} = 0.9625$$

The duration measure for the second forward bucket is

$$\frac{10{,}000 \times 0.9604}{10{,}761.5457} = 0.8924$$

14

Binomial Trees

Learning Objectives

After completing this reading, you should be able to:

- Calculate the value of an American and a European call or put option using a one-step and two-step binomial model.
- Describe how volatility is captured in the binomial model.
- Describe how the value calculated using a binomial model converges as time periods are added.
- Define and calculate delta of a stock option.
- Explain how the binomial model can be altered to price options on stocks with dividends, stock indices, currencies, and futures.

The final three chapters of *Valuation and Risk Models* deal with the valuation of options and other derivatives. This chapter covers a valuation method called *binomial trees*. This technique was proposed by Cox, Ross, and Rubinstein (1979) and is widely used for pricing American-style options and many other derivatives.[1] In the next chapter, we explain the famous Black-Scholes-Merton model. In Chapter 16, we cover the way traders quantify and hedge their exposure to derivatives.

Derivatives are valued using what is called a *no-arbitrage argument*. This means prices are calculated on the assumption that there are no arbitrage opportunities for market participants. Recall that we introduced arbitrage in the context of fixed income markets and the law of one price in Chapter 9. The law of one price states that if portfolios X and Y provide the same cash flows at the same times in the future, they should sell for the same price. Otherwise, a trader can short the more expensive portfolio and buy the cheaper portfolio to lock in a riskless profit. Binomial trees are a convenient way of illustrating how no-arbitrage arguments apply to derivatives.

We will also use binomial trees to introduce what is known as *risk-neutral valuation*. This is a result that allows derivatives to be valued by assuming that market participants require an expected return equal to the risk-free rate on all investments. Risk-neutral valuation is *the* most important principle in derivatives pricing.

14.1 ONE-STEP TREES

Suppose a stock is priced at USD 50 and that it will provide a return of either +20% or −20% over the next three months. In other words, the stock price will either be USD 60 or USD 40 at the end of the three months. While this scenario is not realistic, it provides a useful way of introducing binomial trees. (In this example, we assume no dividends.)

Consider a three-month call option on the stock with a strike price of USD 52. If the stock price turns out to be USD 60, the payoff for the option will be USD 8. If it turns out to be USD 40, the payoff will be zero. The situation is illustrated in Figure 14.1. (We will omit the USD descriptor for the remainder of this section.)

How much is the call option worth? It turns out that the option price can be calculated without any additional information beyond the three-month, risk-free rate. In particular, we do not need to know the probabilities of the two outcomes in Figure 14.1.

[1] See J. C. Cox, S. A. Ross, and M. Rubinstein, "Option Pricing: A Simplified Approach," *Journal of Financial Economics* 7 (October 1979): 229–264.

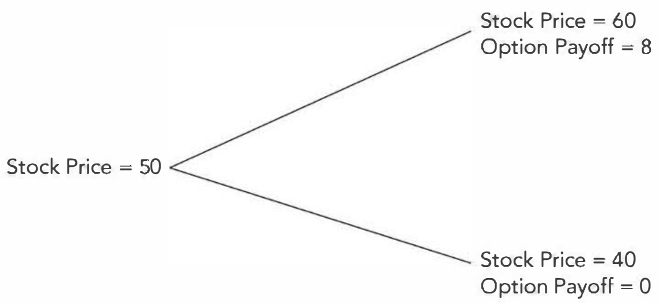


Figure 14.1 Stock price movements and the payoff from a call option with strike price 52.

Consider a portfolio consisting of a long position in Δ shares of the stock and a short position in one call option (Δ is the Greek capital letter for delta). The value of the portfolio in three months will be one of the following.

- If the stock price moves up to 60, the option is worth 8, and the value of the portfolio is $60\Delta - 8$ (note that the option costs the portfolio 8 because the short position will require a stock worth 60 to be sold for 52).
- If the stock price moves down to 40, the option is worth nothing, and the value of the portfolio is 40Δ.

If:

$$60\Delta - 8 = 40\Delta$$

the portfolio then has the same value in three months for both outcomes. Solving this equation, we get $\Delta = 0.4$ and thus a final portfolio value of 16 (= $(60 \times 0.4) - 8$ and 40×0.4).[2]

Suppose next that the three-month, risk-free rate is 3% (with continuous compounding). A portfolio certain to be worth 16 in three months must be worth 15.880 (= $16e^{-0.03 \times 0.25}$) today. Otherwise, there would be arbitrage opportunities. (If the portfolio is worth less than 15.880, an arbitrageur can earn more than 3% by taking a long position in the portfolio. If it is worth more than 15.880, an arbitrageur can borrow at less than 3% by shorting the portfolio.)

Suppose the value of the call option is f. The value of a long position in 0.4 shares is 20 (= 0.4×50). The value of the portfolio today is therefore $20 - f$. For no-arbitrage, we must have

$$20 - f = 15.880$$

[2] Trading 0.4 shares is, of course, not possible. To overcome any concerns on this front, we can multiply the size of the portfolio by 1,000, so that 1,000 options (ten contracts) are sold, and 400 shares are purchased. The resulting price of an option to buy one share is the same. (This assumes that both the option and the stock trade in liquid markets.)

or $f = 4.120$. Our conclusion is that an option to purchase one share must be worth 4.120, as any other price would present arbitrage opportunities.

Generalization

To generalize the argument just illustrated, suppose that the price of a non-dividend paying stock is currently S, and that during a time T it will either move up to Su (providing a return of $u - 1$) or down to Sd (providing a return of $d - 1$). We consider a derivative (which need not be an option) that provides a payoff of f_u if the stock price increases, and a payoff of f_d if the stock price decreases. The situation is illustrated in Figure 14.2.

We then form a portfolio consisting of:

- A short position in one unit of the derivative; and
- A position of Δ in the stock

where:

$$\Delta = \frac{f_u - f_d}{Su - Sd} \tag{14.1}$$

The value of the portfolio at time T is

- $Su\Delta - f_u$ if the stock price increases, and
- $Sd\Delta - f_d$ if the stock price decreases.

These two are the same for the value of Δ in Equation (14.1). For this value of Δ, the value of the portfolio at time T (for both upward and downward movements in the stock price) is

$$Su\frac{f_u - f_d}{Su - Sd} - f_u = \frac{f_u d - f_d u}{u - d} = Sd\frac{f_u - f_d}{Su - Sd} - f_d$$

The value of the portfolio today is $S\Delta - f$, where f is the value of the derivative today. Suppose r is the risk-free rate for maturity T. For no arbitrage, we must have

$$S\Delta - f = \frac{f_u d - f_d u}{u - d} e^{-rT}$$

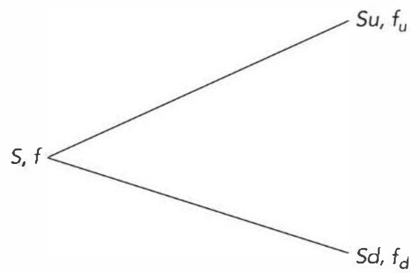


Figure 14.2 Generalized one-step model. S and f are the initial stock price and derivative price, and f_u and f_d are the payoffs from the derivative when there are up and down stock price movements, respectively.

Substituting for Δ from Equation (14.1) gives

$$f_u - f_d - f(u - d) = (f_u d - f_d u)e^{-rT}$$

or

$$f = f_u \frac{1 - de^{-rT}}{u - d} + f_d \frac{ue^{-rT} - 1}{u - d}$$

This can be written as:

$$f = e^{-rT}[pf_u + (1 - p)f_d] \tag{14.2}$$

where:

$$p = \frac{e^{rT} - d}{u - d} \tag{14.3}$$

As an illustration of Equations (14.2) and (14.3), suppose the option in Figure 14.2 is a put option with strike price 52 (instead of a call option with that strike price). Equation (14.3) gives

$$p = \frac{e^{0.03\times0.25} - 0.8}{1.2 - 0.8} = 0.5188$$

In this case, $f_u = 0$ and $f_d = 12$ so that Equation (14.2) gives the value of the put option as:

$$e^{-0.03\times0.25}(0.5188 \times 0 + 0.4812 \times 12) = 5.731$$

We introduced put-call parity when discussing the properties of stock options in Financial Markets and Products. This states that when a European put option and a European call option on a non-dividend paying stock have the same strike price and time to maturity:

Call Price + PV of Strike Price = Put Price + Stock Price

Put-call parity holds for our example. The value of the call calculated earlier is 4.120, the present value of the strike price is 51.611 ($= 52e^{-0.03\times0.25}$), the put price we have just calculated is 5.731, and the stock price is 50. We therefore have

$$4.120 + 51.611 = 5.731 + 50$$

It is important to note the analysis in this section does not require us to know the probabilities on the up and down branches. This is because we are valuing the derivative in terms of the stock price, and therefore the return from the derivative depends on the return on the stock. It turns out the relationship between the two returns is such that we do not need to know either return to do the analysis.

14.2 RISK-NEUTRAL VALUATION

Equations (14.2) and (14.3) illustrate an important principle in option pricing known as *risk-neutral valuation*.

We define a risk-neutral world as one where investors do not adjust their required expected returns for risk, so that the expected return on all assets is the risk-free rate. To put this another way, a risk-neutral world is one where all tradable assets have an expected return equal to the risk-free interest rate. The probabilities of different outcomes in a risk-neutral world are therefore based on this assumption, and a risk-neutral investor has no preference between assets with different risks.

The risk-neutral valuation principle states that if we assume we are in a risk-neutral world, we get the correct price for a derivative. As it turns out, the price is correct in the real world (where investors do care about risk) as well as in a risk-neutral world.

First, we note that if we choose to interpret the variable p in Equation (14.2) as the probability of an upward movement (with $1 - p$ being the probability of a downward movement), then:

$$pf_u + (1 - p)f_d$$

is the expected payoff from the derivative at time T. Meanwhile, the value of the derivative (f) is the present value of the expected payoff at time T with the discount rate being the risk-free rate.

Second, we note that if we continue to interpret p as the probability of an upward movement, the expected stock price is

$$Sup + Sd(1 - p)$$

Substituting for p from Equation (14.3), this becomes

$$Su\frac{e^{rT} - d}{u - d} + Sd\frac{u - e^{rT}}{u - d} = Se^{rT}$$

This shows that the stock price grows at the risk-free rate. It also means that p is the probability of an upward movement in a risk-neutral world.

We have therefore demonstrated the truth of the risk-neutral valuation result. If we assume a risk-neutral world, the probability of an upward movement is p, and the value of the derivative is its expected payoff discounted at the risk-free rate (i.e., it is the value that would apply if market participants were risk neutral).

It should be emphasized that the risk-neutral valuation is nothing more than an artificial way of valuing derivatives. We are not assuming that the world is actually risk-neutral. We are instead arguing that the price of a derivative is the same in the real world as it would be in the risk-neutral world.

Suppose that the probability of an upward movement in Figure 14.1 in the real world is 0.6. This means the expected return on the stock over three months is 4% (= 0.6 × 20% − 0.4 × 20%). The expected return on the call option over three months in the real world is

$$\frac{0.6 \times 8 + 0.4 \times 0}{4.120} - 1 = 16.5\%$$

This shows the expected return from the call option is much higher than that for the stock in the real world we are assuming. The stock provides a return that reflects its systematic risk. The call option has built-in leverage and therefore accentuates this risk.

Now suppose the option is a put option with a strike price of 52 (rather than a call option). As we saw in the previous section, the value of this option is 5.731. The expected return from the option during the three months is

$$\frac{0.6 \times 0 + 0.4 \times 12}{5.731} - 1 = -16.2\%$$

This may seem to be a surprising result. Why does a risky security, such as a put option, provide a negative expected return? The answer is that the stock price provides a greater return than the risk-free rate because it has positive systematic risk (when the market does well, the stock does well; when the market does poorly, the stock does poorly). The return from the put option is negatively related to the return from the stock, and it therefore has negative systematic risk. This means that its expected return is less than the risk-free rate. (The leverage in a put option increases the amount by which the put option's expected return is less than the risk-free rate.)

14.3 MULTI-STEP TREES

Of course, it is quite unrealistic to model stock price changes using a tree with a single step (as in Figures 14.1 and 14.2). To create a more realistic valuation model, the life of an option is divided into many steps that can be handled in the same way as the steps in Figures 14.1 and 14.2.

As is the normal practice, we will now define the length of a tree step as Δt. This means we can replace T with Δt in Equations (14.2) and (14.3) so that:

$$f = e^{-r\Delta t}[pf_u + (1 - p)f_d] \quad (14.4)$$

where:

$$p = \frac{e^{r\Delta t} - d}{u - d} \quad (14.5)$$

Before moving on to show how these equations are used in multi-step trees, we will explain how u and d are determined.

The parameters u and d should be chosen to reflect the volatility of the stock price.[3] If we denote the volatility per year by σ, then appropriate values for the parameters are

$$u = e^{\sigma\sqrt{\Delta t}}$$
$$d = e^{-\sigma\sqrt{\Delta t}} \quad (14.6)$$

where Δt is measured in years. The appendix to this chapter explains why these parameters match the volatility.

[3] See Section 3.3 for a discussion of the measurement of volatility.

A Two-Step Tree

Consider the situation where the price of a non-dividend paying stock is 29 with a volatility of 25%. We will consider a European call option with a strike price of 30 and suppose the life of the option is one year. Finally, we will assume that the risk-free rate is 3% (continuously compounded).

In this case, $\Delta t = 0.5$. Using Equation (14.6):

$$u = e^{0.25\sqrt{0.5}} = 1.1934$$

$$d = e^{-0.25\sqrt{0.5}} = 0.8380$$

To reflect the volatility, the upward movement in the stock price provides a return of 19.34%, and the downward movement provides a return of −16.20%.

From Equation (14.5):

$$p = \frac{e^{0.03\times0.5} - 0.8380}{1.1934 - 0.8380} = 0.4984$$

This means that in a risk-neutral world, the probability of an upward movement should be assumed to be 0.4984, and therefore the probability of a downward movement is 0.5016.

Figure 14.3 shows the tree.[4] The upper number at each node is the stock price, and the lower number is the option price. We first calculate the stock prices at each node. The stock price at node B is 34.608 ($= 29u = 29 \times 1.1934$), the stock price at node D is 41.299 ($= 34.608u$), and the stock price at node C is 24.301 ($= 29d = 29 \times 0.8380$).[5] Note that the tree recombines: Moving from A to B to E leads to a stock price of 29ud at E, while moving from A to C to E leads to a stock price of 29du.

We value the option by starting at the end of the tree and working back. This is referred to as *backward induction* (or rolling back through the tree). The value of the option at node D is the payoff if node D is reached. It is 11.299 ($= 41.299 - 30$). If either node E or F is reached, the option should not be exercised and so there is zero payoff.

Consider next the value of the option at node B. Using Equation (14.4) this is

$$e^{-0.03\times0.5}(0.4984 \times 11.299 + 0.5016 \times 0) = 5.548$$

[4] This tree and others in the rest of this chapter are from the RMFI software that accompanies John Hull's book, *Risk Management and Financial Institutions*, 5th edition, Wiley. The software can be downloaded from http://www-2.rotman.utoronto.ca/~hull/riskman

[5] The values given here and elsewhere are calculated by retaining more decimal places than those shown.

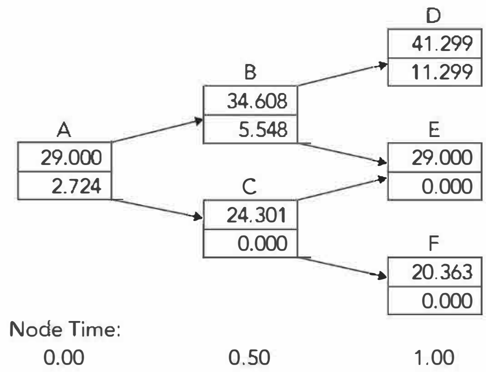


Figure 14.3 **Shown is a two-step tree for valuing a call option with a strike price of 30. The upper number at each node is the stock price, and the lower number is the option price.**

The value of the option at node C is zero because it is worth zero at the two nodes that can be reached from C. The value at node A is

$$e^{-0.03\times0.5}(0.4984 \times 5.548 + 0.5016 \times 0) = 2.724$$

We can therefore deduce that the two-step tree gives the value of the option today as 2.724.

A Put Example

We now change the example in Figure 14.3 to be a put option (rather than a call option). This leads to Figure 14.4. The stock prices on the tree are the same, but the option prices are different. The option is exercised at nodes E and F. At node E, there is a payoff of 1, while the payoff at node F is 9.673. There is no payoff at node D because the stock price is above the strike price of 30.

The procedure for rolling back through the tree is the same as before. For example, the value at node B is

$$e^{-0.03\times0.5}(0.4984 \times 0 + 0.5016 \times 1) = 0.494$$

The value at node C is

$$e^{-0.03\times0.5}(0.4984 \times 1.000 + 0.5016 \times 9.637) = 5.252$$

The value at node A is

$$e^{-0.03\times0.5}(0.4984 \times 0.494 + 0.5016 \times 5.252) = 2.838$$

This is the current value of the option given by the tree.

American Options

Up to now, we have only considered European options (i.e., options that can only be exercised at maturity). We now consider how our calculations change when we are dealing with American options (i.e., options that can be exercised at any time).

At each node, we must carry out two calculations to determine:

- How much the option is worth if it is exercised at the node, and
- How much the option is worth if it is not exercised.

The value at the node is the greater of these.

Consider again the put option in Figure 14.4. However, now suppose that it is American. This leads to Figure 14.5. The values at the final nodes are the same as in Figure 14.4. If those nodes are reached (and they may not be reached if the option is exercised early), the value of the option has its intrinsic value of $\max(K - S, 0)$ where K is the strike price, and S is the stock price.

Consider node B. If the option is exercised at node B, it gives a payoff of -4.608 (i.e., a loss of 4.608). Clearly it should not be exercised, and thus the value at the node is 0.494 (as in Figure 14.4). At node C, we have two values.

1. The value of the option if it is not exercised is 5.252 (as in Figure 14.4).
2. The value of the option if it is exercised is 5.699 ($= 30 - 24.301$).

Since $5.699 > 5.252$, the option should be exercised at node C, and thus the value at this node is 5.699.

At node A, the value of the option if it is exercised is 1.000. The value if it is not exercised is

$$e^{-0.03\times 0.5}(0.4984 \times 0.494 + 0.5016 \times 5.699) = 3.058$$

It is therefore not worth exercising the option at node A, and thus the initial value of the option given by the two-step tree is 3.058. We saw earlier that the European option is worth 2.838. In this case, the value of the American feature is therefore

$$3.058 - 2.838 = 0.22$$

Note that the roll back procedure means that the value we calculate at a node reflects not only the possibility of immediate early exercise, but also the possibility of exercise at later nodes.

Increasing the Number of Steps

Trees with more than two steps can be handled like the trees in Figures 14.3 to 14.5. Suppose that we use four steps rather than two steps when valuing the American option in Figure 14.5. We then have $\Delta t = 0.25$ and:

$$u = e^{0.25\sqrt{0.25}} = 1.1331$$

$$d = e^{-0.25\sqrt{0.25}} = 0.8825$$

$$p = \frac{e^{0.03\times 0.25} - 0.8825}{1.1331 - 0.8825} = 0.4988$$

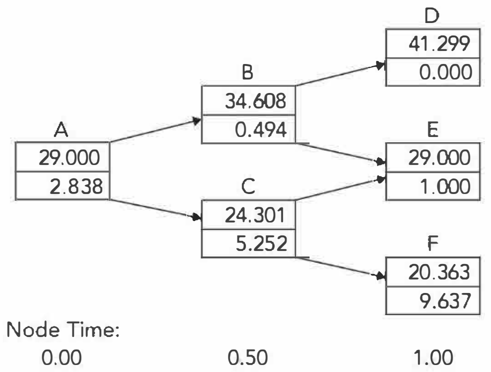


Figure 14.4 Shown is a two-step tree for valuing a put option with a strike price of 30. The upper number at each node is the stock price, and the lower number is the option price.

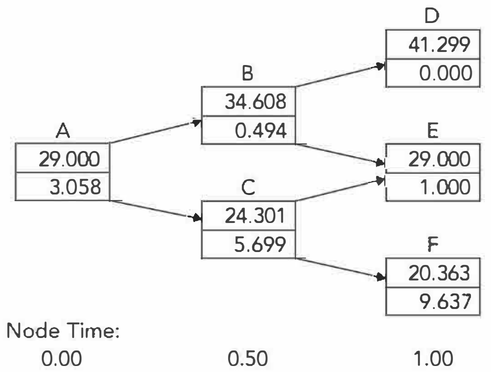


Figure 14.5 Shown is a two-step tree for valuing an American put option with a strike price of 30. The upper number at each node is the stock price, and the lower number is the option price.

The tree is shown in Figure 14.6. The option is exercised early at nodes F, I, and J. The value calculated for the option is 3.097, which is a little higher than that given in Figure 14.5 using two steps.

The accuracy increases as the number of steps increases. For the option we are considering, 20, 50, 100, and 500 steps give values of 3.082, 3.067, 3.059, and 3.055 (respectively). In practice, at least 30 to 50 steps are usually used. In the next chapter, we will present the Black-Scholes-Merton formula for valuing European options and explain the underlying random-walk assumption. The binomial tree makes the same random-walk assumption, and it can be shown that, as the number of steps is increased, the binomial tree valuation of a European option converges to its Black-Scholes-Merton valuation.

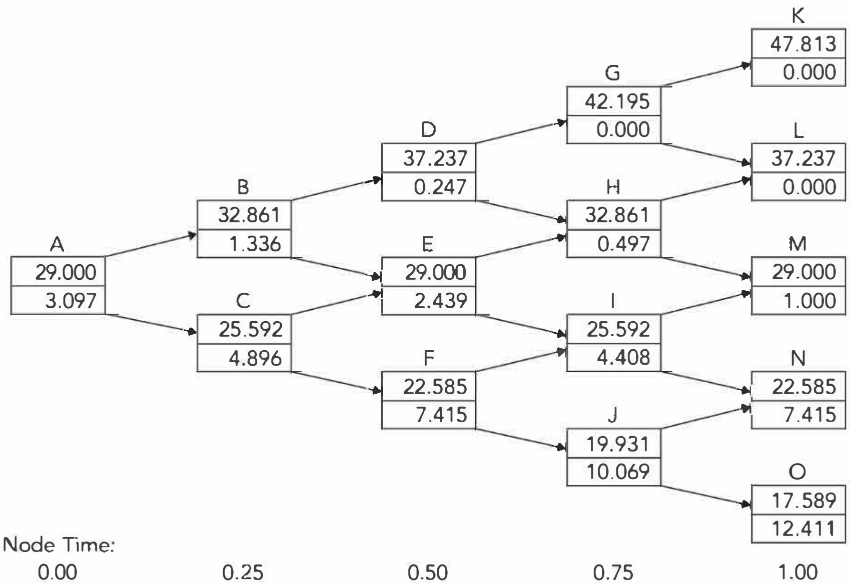


Figure 14.6 Shown is a four-step tree for valuing an American put option with a strike price of 30. The upper number at each node is the stock price, and the lower number is the option price.

14.4 DELTA

We now return to Equation (14.1).

$$\Delta = \frac{f_u - f_d}{Su - Sd}$$

Recall that Δ is the position taken in the stock to hedge a short position in one unit of the derivative. This is an important hedge parameter (referred to as a Greek letter) and will be discussed further in Chapter 16.

Delta is the sensitivity of a derivative's value to the price of its underlying stock. When the stock price changes from Sd to Su, the option price changes from f_d to f_u. (Theoretically delta is a measure of the sensitivity at time Δt, not at time zero. In a large tree where Δt is extremely small, however, there is very little difference between the delta at time zero and the delta at time Δt.)

Consider the tree in Figure 14.6. We would calculate delta as:

$$\frac{1.336 - 4.896}{32.861 - 25.592} = -0.490$$

This is negative because we are dealing with a put option whose price is negatively related to the underlying asset's price. This indicates that when the stock price increases (or decreases) by a small amount, the put option price decreases (or increases) by about 49% of that amount.

Figure 14.6 illustrates that delta does not remain constant through time. For example, the delta calculated from the nodes D and E at time $2\Delta t$ is

$$\frac{0.247 - 2.439}{37.237 - 29.000} = -0.266$$

On the other hand, the delta calculated from nodes E and F at time $2\Delta t$ is

$$\frac{2.439 - 7.415}{29.000 - 22.585} = -0.776$$

14.5 OTHER ASSETS

So far, we have only considered options on non-dividend paying stocks. To consider other assets, it will prove useful to consider a stock paying a continuous dividend yield at rate q. (This means the dividend paid during a short time period Δt is the stock price multiplied by q multiplied by Δt.)[6]

A dividend yield of q means that the formulas we have presented so far must be adjusted slightly. The total return in a risk-neutral world is r. Dividends provide a return of q. The expected growth of the stock price must therefore be $r - q$. Recall that we have

$$pSu + (1 - p)Sd = Se^{r\Delta t}$$

when there are no dividends so that:

$$p = \frac{e^{r\Delta t} - d}{u - d}$$

In the case where dividends are paid at rate q, we have

$$pSu + (1 - p)Sd = Se^{(r-q)\Delta t}$$

so that:

$$p = \frac{e^{(r-q)\Delta t} - d}{u - d} \tag{14.7}$$

Everything else about the tree, including the calculation of u and d and the roll back procedure, is the same as before.

Stock Indices

For an option on a stock index, it is usually assumed the index provides a dividend yield. The valuation of an option on a stock index should therefore involve the modification in

[6] In practice, stocks pay discrete rather than continuous dividends, but, as we will show, some other assets that underlie derivatives are analogous to stocks paying continuous dividends.

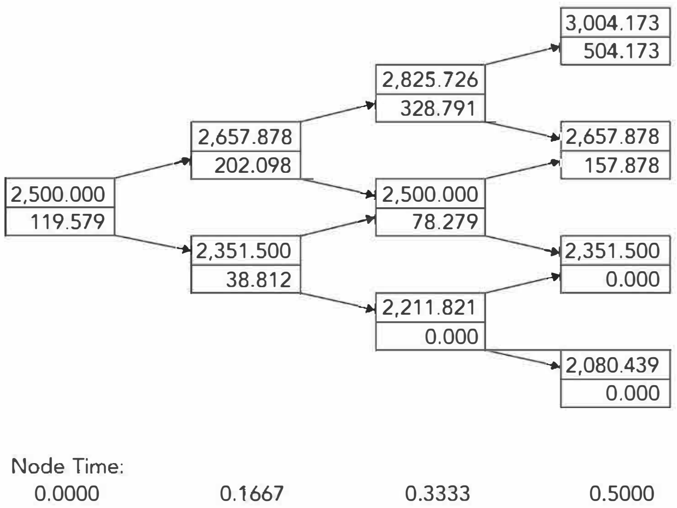


Figure 14.7 Shown is a three-step tree for valuing a European call option on an index with a strike price of 2,500. The upper number at each node is the stock index price, and the lower number is the option price.

Equation (14.7). We set q equal to the estimated average dividend yield during the life of the option.

Consider a stock index standing at 2,500. Suppose the dividend yield on the index is 2% while the risk-free rate is 3%. Suppose further that the volatility of the index is 15% per annum. Figure 14.7 uses a three-step tree to value a European call option with a strike price of 2,500 and a time to maturity of six months.

In this case:

$$u = e^{0.15\sqrt{0.1667}} = 1.0632$$

$$d = e^{-0.15\sqrt{0.1667}} = 0.9406$$

$$p = \frac{e^{(0.03-0.02)\times 0.1667} - 0.9406}{1.0632 - 0.9406} = 0.4983$$

and the value of the option given by the three-step tree is 119.579.

Currency

As noted in Financial Markets and Products, a currency can be considered as an asset providing a yield at the foreign risk-free rate. Therefore, the analysis we presented for a stock paying a continuous dividend yield applies, with q equal to the foreign risk-free rate (r_f). This means that:

$$p = \frac{e^{(r-r_f)\Delta t} - d}{u - d}$$

The rest of the analysis is as explained earlier in this chapter for non-dividend paying stocks.

As an example, consider a fourstep tree for a one-year American option to buy a foreign currency for 0.8000 when the current exchange rate is 0.7800. The volatility of the exchange rate is 12%, while the domestic and foreign risk-free rates are 2% and 6% (respectively). In this case:

$$u = e^{0.12\sqrt{0.25}} = 1.0618$$

$$d = e^{-0.12\sqrt{0.25}} = 0.9418$$

$$p = \frac{e^{(0.02-0.06)\times 0.25} - 0.9418}{1.0618 - 0.9418} = 0.4021$$

The tree is shown in Figure 14.8. (The option is exercised early at nodes A and B. The value of the option to buy one unit of the foreign currency is 0.0188.)

Futures

Because it costs nothing to enter into a futures contract, the return on a futures contract in a risk-neutral world must be zero. This means we can treat a futures contract like a stock, paying a continuous dividend yield equal to r. This is because when $q = r$, the expected growth rate of the stock is zero.

From Equation (14.7), we therefore set

$$p = \frac{1 - d}{u - d}$$

and otherwise proceed in the same way as for a non-dividend paying stock.

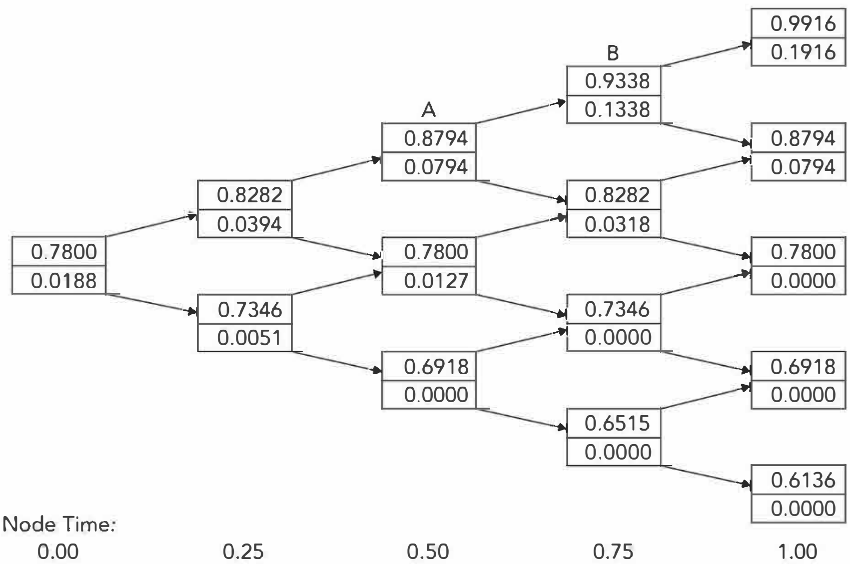


Figure 14.8 Shown is a four-step tree for valuing an American call option on a currency with a strike price of 0.8000. The upper number at each node is the currency exchange rate, and the lower number is the option price.

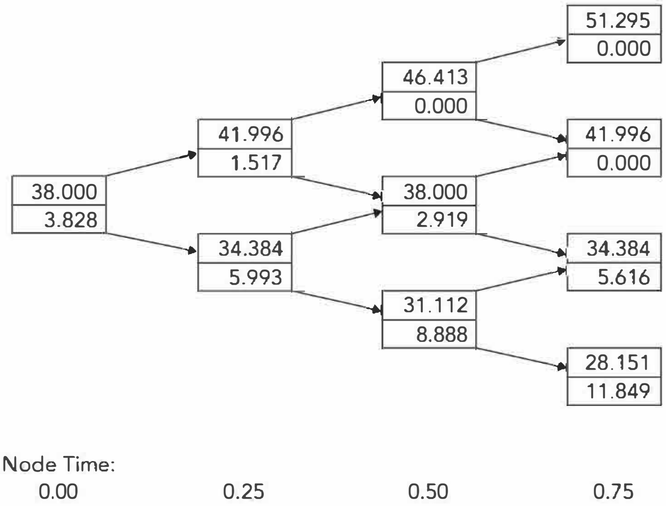


Figure 14.9 Shown is a three-step tree for valuing an American put option on a futures contract with a strike price of 40. The upper number at each node is the futures price, and the lower number is the option price.

For example, consider a three-step tree to value an American nine-month put option on a futures contract when the current futures price is 38, and the strike price is 40. We assume the volatility is 20%, and the risk-free rate is 4%. The tree is shown in Figure 14.9. In this case:

$$u = e^{0.2\times\sqrt{0.25}} = 1.1052$$

$$d = e^{-0.2\times\sqrt{0.25}} = 0.9048$$

$$p = \frac{1 - 0.9048}{1.1052 - 0.9048} = 0.4750$$

The value of the option is 3.828.

SUMMARY

In the simple (but unrealistic) situation where the movements in a stock price can be represented by a one-step binomial tree, it is possible to set up a portfolio consisting of an option and the stock that is riskless. Because riskless portfolios must earn the risk-free rate, this enables the option to be valued in terms of the stock's price.

When stock price movements are governed by a multi-step tree, we can treat each binomial step separately and roll back through the tree to value a derivative. For American options, it is necessary to test for early exercise at each node of the tree.

A very important general principle in options pricing states that, when valuing derivatives, we can assume market participants are risk-neutral so that their expected return is the risk-free rate. We then get the correct valuation for a derivative in the real world (as well as in the risk-neutral world). Binomial trees provide an illustration of this important principle.

Trees can be constructed for valuing derivatives dependent on stock indices, currencies, and futures in the same way that they are used to value derivatives dependent on a non-dividend paying stock. All three are analogous to a stock paying a continuous dividend yield and can be handled by making an adjustment to the risk-neutral probabilities on the tree. In the case of the stock index, the dividend yield is the dividend yield on the index; in the case of a currency, it is the foreign risk-free rate; in the case of a futures contract, it is the domestic risk-free rate.

APPENDIX

The Formulas for *u* and *d*

The variance of the return on a stock in a short period of time (Δt) is $\sigma^2\Delta t$, where σ is the volatility. Using the standard formula from statistics, the variance of the return is $E(R^2) - [E(R)]^2$ where R is the return, and E denotes the expected value. In a step of a binomial tree, the return has a probability of p of being $u - 1$ and a probability $1 - p$ of being $d - 1$. Hence, we are required to choose u and d so that:

$$p(u-1)^2 + (1-p)(d-1)^2 - [p(u-1) + (1-p)(d-1)]^2 = \sigma^2\Delta t$$

Substituting for p from Equation (14.3), this becomes (after some simplification):

$$e^{r\Delta t}(u + d) - ud - e^{2r\Delta t} = \sigma^2\Delta t \qquad (14A.1)$$

When terms of an order higher than Δt are ignored in the expansion of the exponential function, we obtain

$$e^{\sigma\sqrt{\Delta t}} = 1 + \sigma\sqrt{\Delta t} + \frac{1}{2}\sigma^2\Delta t$$

$$e^{-\sigma\sqrt{\Delta t}} = 1 - \sigma\sqrt{\Delta t} + \frac{1}{2}\sigma^2\Delta t$$

$$e^{r\Delta t} = 1 + r\Delta t$$

$$e^{2r\Delta t} = 1 + 2r\Delta t$$

It can be seen that $u = e^{\sigma\sqrt{\Delta t}}$ and $d = e^{-\sigma\sqrt{\Delta t}}$ satisfy Equation (14A.1).

A final point to note is that a result known as Girsanov's theorem states that volatility is the same in the risk-neutral world as it is in the real world.[7] This is convenient as it enables volatility to be estimated from historical data and used in risk-neutral valuation calculations.

[7] *See* I. V. Girsanov, "On transforming a certain class of stochastic processes by absolutely continuous substitution of measures," Theory of Probability and its Applications, 5, 3 (1960): 285–301.

The following questions are intended to help candidates understand the material. They are <u>not</u> actual FRM exam questions.

QUESTIONS

Short Concept Questions

14.1 What is a no-arbitrage argument for pricing?

14.2 What is the principle known as risk-neutral valuation?

14.3 What are the Cox, Ross, Rubinstein formulas for u and d in terms of the volatility, σ, and the length of the time step, Δt?

14.4 What is the formula for the risk-neutral probability of an upward movement in the case of a stock paying no dividend?

14.5 How does the formula in Question 14.4 change for (a) a stock index, (b) a currency, and (c) a futures price?

14.6 How many steps are typically used to value an option using the binomial tree methodology?

14.7 What is the difference between using a binomial tree to value an American option and using it to value the corresponding European option?

14.8 How is the delta of an option defined?

14.9 How can delta be estimated from a binomial tree?

14.10 In what way is a futures price analogous to a stock price paying a continuous dividend yield?

Practice Questions

14.11 A stock price is currently 40. It is known that it will be 42 or 38 at the end of a month. The risk-free rate is 4% per annum with continuous compounding. What is the value of a one-month call option with a strike price of 39?

14.12 In Question 14.11, what position should be taken in the stock to hedge a short position in the option?

14.13 A stock price is currently 40. At the end of six months it will be either 36 or 44. The risk-free rate is 5% per annum with continuous compounding. What is the value of a six-month European put option with a strike price of 40?

14.14 In Question 14.13, what position should be taken in the stock to hedge a long position in the option?

14.15 A stock price is currently 50. Its volatility is 20% per annum. The risk-free rate is 4% per annum with continuous compounding. Use a two-step tree to determine the value of a six-month European call option on the stock with a strike price of 48.

14.16 In Question 14.15, what is the value of a European put option with a strike price of 48? Check that put-call parity holds.

14.17 In Question 14.15, value an option that pays off $\max(S^2 - 2{,}400, 0)$ in six months where S is the stock price. (This is known as a power option.)

14.18 Use a two-step tree to value a one-year American call option on an index. The current value of the index is 2,000, the risk-free rate is 2%, and the dividend yield on the index is 3%. The strike price is 1,900 and the volatility is 22% per annum.

14.19 Use a two-step tree to value a six-month American put option on a foreign currency for a US investor. The current value of the currency is USD 1.3000, the US risk-free rate is 3%, and the foreign risk-free rate is 5%. The strike price is 1.3200, and the volatility is 14% per annum.

14.20 Use a two-step tree to value an eight-month American put option on a futures contract. The current futures price is 58 and the risk-free rate is 5%. The strike price is 60 and the volatility is 24% per annum.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

ANSWERS

Short Concept Questions

14.1 A no-arbitrage argument is an argument that prices should adjust so that there are no arbitrage opportunities.

14.2 Risk-neutral valuation states that we can price derivatives on the assumption that the world is risk-neutral (i.e., market participants do not adjust their required return for the risk they are taking). The price we get is correct in the real world as well as in the risk-neutral world that is assumed.

14.3 $u = e^{\sigma\sqrt{\Delta t}}$ and $d = e^{-\sigma\sqrt{\Delta t}}$

14.4 It is

$$p = \frac{e^{r\Delta t} - d}{u - d}$$

14.5 (a) Replace $e^{r\Delta t}$ by $e^{(r-q)\Delta t}$ where q is the dividend yield on the index.

(b) Replace $e^{r\Delta t}$ by $e^{(r-r_f)\Delta t}$ where r_f is the foreign risk-free rate.

(c) Replace $e^{r\Delta t}$ by 1.

14.6 30–50 steps

14.7 In the case of an American option, we have to check at each node whether the option should be exercised. The value at a node is the greater of the intrinsic value of the option and the result of rolling back from the subsequent nodes.

14.8 The delta of an option is the sensitivity of its price to the price of the underlying asset. It is $\Delta f/\Delta S$ where f is the option price and S is the price of the underlying asset.

14.9 Delta can be calculated from the two nodes at time Δt. It is the increase in the option price when one moves from the lower node to the upper node divided by the increase in the stock price when one does so.

14.10 It costs nothing to enter into a futures contract. A futures price should therefore have zero growth rate in a risk-neutral world. The same is true of a stock that provides a dividend yield equal to the risk-free rate.

Solved Problems

14.11 In this case, $u = 42/40 = 1.05$ and $d = 38/40 = 0.95$ so that:

$$p = \frac{e^{0.04\times0.0833} - 0.95}{1.05 - 0.95} = 0.5334$$

and the value of the option is

$$(0.5334 \times 3 + 0.4666 \times 0) \times e^{-0.04\times0.0833} = 1.595$$

14.12 The position is long 0.75 of a share. This is because a portfolio of 0.75 shares and short one option is worth 28.5 for both outcomes.

14.13 In this case, $u = 44/40 = 1.1$ and $d = 36/40 = 0.9$ so that:

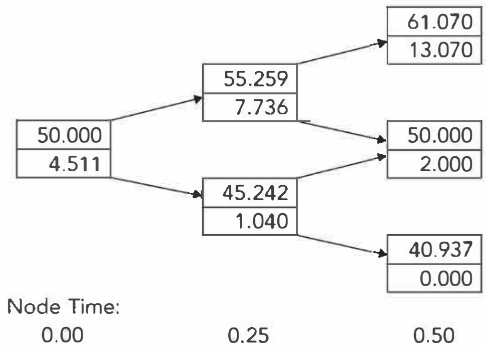
$$p = \frac{e^{0.05\times0.5} - 0.9}{1.1 - 0.9} = 0.6266$$

and the value of the option is

$$(0.6266 \times 0 + 0.3734 \times 4) \times e^{-0.05\times0.5} = 1.4568$$

14.14 The position is long 0.5 shares. This is because 0.5 shares plus the long put option is worth 22 for both outcomes.

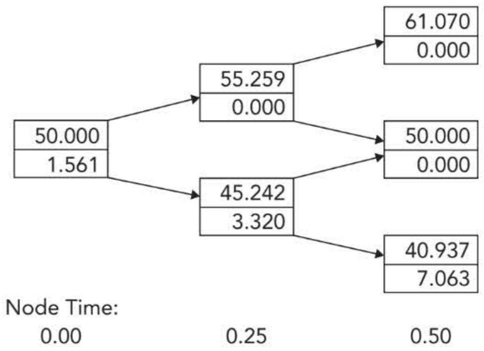
14.15 In this case, $u = 1.1052$, $d = 0.9048$, and $p = 0.5252$. The following two-step tree shows that the value of the option is 4.511.



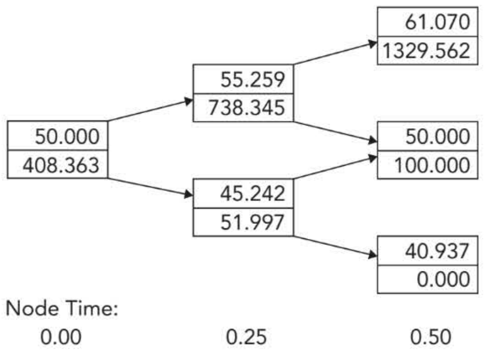
The following questions are intended to help candidates understand the material. They are *not* actual FRM exam questions.

14.16 As the following tree shows, the value of the put option is 1.561. Put-call parity holds because:

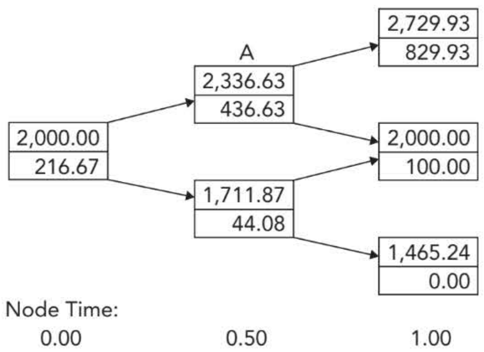
$$4.511 + 48e^{-0.04\times0.5} = 51.561 = 50 + 1.561$$



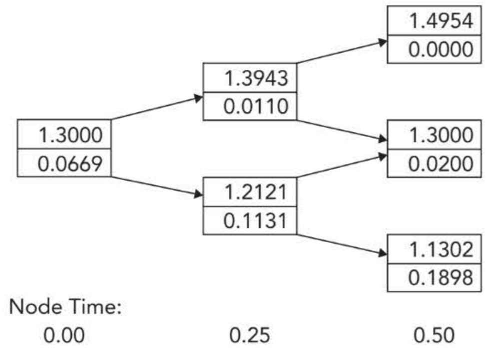
14.17 For the power option, the tree becomes as follows, and the value of the option is 408.363.



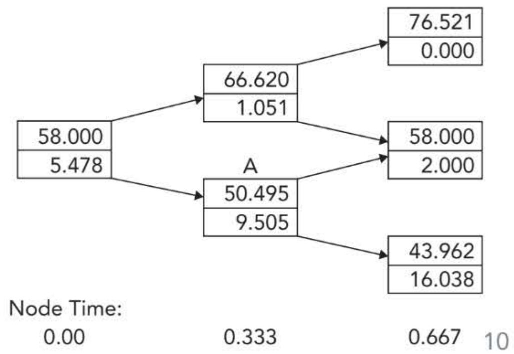
14.18 The tree is shown as follows. The option is exercised early at node A. The value of the option is 216.67.



14.19 In this case, there is no early exercise. The value of the option per unit of foreign currency is 0.0669.



14.20 The option is exercised at node A. The value today is 5.478.



15

The Black-Scholes-Merton Model

Learning Objectives

After completing this reading, you should be able to:

- Explain the lognormal property of stock prices, the distribution of rates of return, and the calculation of expected return.
- Calculate the realized return and historical volatility of a stock.
- Describe the assumptions underlying the Black-Scholes-Merton option pricing model.
- Calculate the value of a European option on a non-dividend-paying stock using the Black-Scholes-Merton model.
- Define implied volatilities and describe how to calculate implied volatilities from market prices of options using the Black-Scholes-Merton model.
- Explain how dividends affect the decision to exercise early for American call and put options.
- Calculate the value of a European option on a dividend-paying stock, futures, or foreign currency using the Black-Scholes-Merton model.
- Describe warrants, calculate the value of a warrant, and calculate the dilution cost of the warrant to existing shareholders.

Financial Markets and Products introduced options, explained some of their properties, and discussed trading strategies. We now explain the famous Black-Scholes-Merton model. This model was published in two papers in 1973 and has had a major influence on the way options are priced and hedged.[1] In one of the papers, Black and Scholes used the capital asset pricing model (CAPM) to derive the relationship between the return from a stock and the return from an option on the stock. In the other paper, Merton used no-arbitrage arguments like those used in connection with binomial trees in the previous chapter. The two papers derived the same option pricing formula.

The pricing formula applies to European options on non-dividend paying stocks. As we will show, it can be extended to European options on stocks paying discrete dividends and to European options on other assets (such as stock indices, currencies, and futures). It does not apply to American options, which must be valued using the binomial tree methodology explained in the previous chapter.

15.1 STOCK PRICE MOVEMENTS

The Black-Scholes-Merton model assumes that the return from a non-dividend paying stock over a short period of time is normally distributed. If μ is the mean return and σ is the volatility, then the return in time Δt is assumed to be normal with mean $\mu\Delta t$ and standard deviation $\sigma\sqrt{\Delta t}$. In theory, we only assume this is true in the limit as Δt tends to zero. In practice, it can be assumed to be approximately true for a small Δt.

Suppose a stock price is USD 100, the mean return (μ) is 12% per year, and the volatility is 25% per year. The probability distribution for the stock's return over one week is approximately normal with mean:

$$\frac{12\%}{52} = 0.23\%$$

and standard deviation:

$$25\% \times \sqrt{\frac{1}{52}} = 3.47\%$$

A 95% confidence interval for the return is

$$0.23\% \pm 1.96 \times 3.47\%$$

or from about −6.6% to +7.0%.[2] This means we are 95% certain the stock price at the end of a week will be between USD 93.4 and USD 107.

The Lognormal Distribution

When the return on a stock over a short period is normally distributed, the stock price at the end of a relatively long period has a lognormal distribution. This means the logarithm of the stock price (and not the stock price itself) is normally distributed.

Figures 15.1 and 15.2 compare a normal distribution to a lognormal distribution. The key differences are as follows.

- A normal distribution is symmetrical and the variable can take any value from negative infinity to infinity.
- A lognormal distribution is skewed and the variable can take any positive value.

A future stock price cannot be negative. Our model is consistent with this observation since it leads to a lognormal distribution for future stock prices. Note that the standard deviation of the change in the stock price S in time Δt is $S\sigma\sqrt{\Delta t}$. This standard deviation declines as the stock price declines, with the result being that movements in the stock price become smaller (and thus the stock price cannot become negative).

Define S_0 as the stock price at time zero and S_T as the stock price at time T. It can be shown that:

- The expected value of S_T is given by:

$$E(S_T) = S_0 e^{\mu T} \tag{15.1}$$

- The expected value of the logarithm of the stock price at time T is

$$E(ln(S_T)) = ln(S_0) + \left(\mu - \frac{\sigma^2}{2}\right)T \tag{15.2}$$

- The standard deviation of the logarithm of the stock price is

$$SD(ln(S_T)) = \sigma\sqrt{T} \tag{15.3}$$

Note from Equations (15.1) and (15.2) that the mean of the stock price's logarithm is not the logarithm of the mean stock price:[3]

$$E(ln(S_T)) \neq ln(E(S_T)) \tag{15.4}$$

This is because ln is a non-linear function.

[1] See F. Black and M. Scholes, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81 (May/June 1973): 637–59, and R. C. Merton, "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science* 4 (Spring 1973): 141–183. Myron Scholes and Robert Merton won the Nobel prize for economics for the development of the Black-Scholes model in 1997. Sadly, Fischer Black died in 1995, otherwise, he would also have been a recipient of this prestigious award.

[2] Traders often assume that the return is zero when translating volatilities into confidence intervals for the return over short periods. The same assumption is often made when VaR and ES are calculated (see Chapters 1 to 3).

[3] $ln(E(S_T)) = ln(S_0 e^{\mu T}) = ln(S_0) + \mu T$ which, from Equation (15.2), is greater than $E(ln(S_T))$.

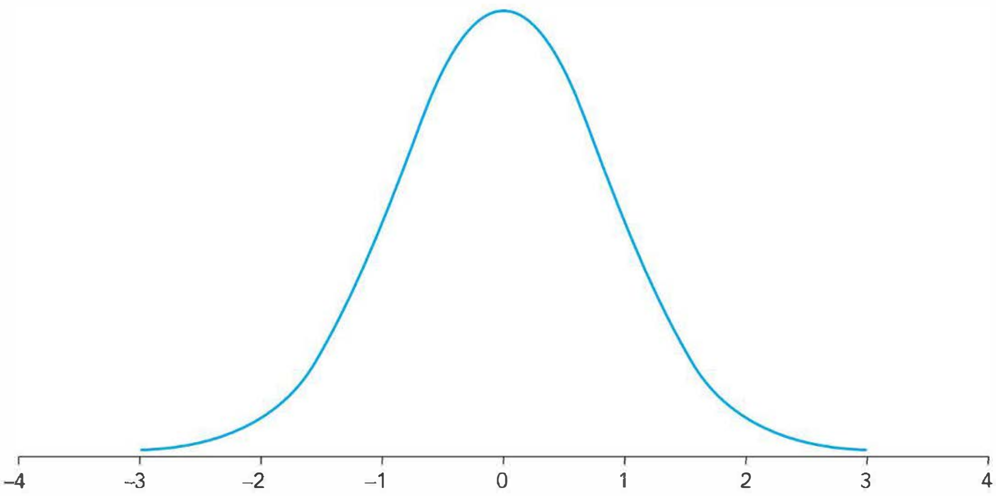


Figure 15.1 Normal distribution.

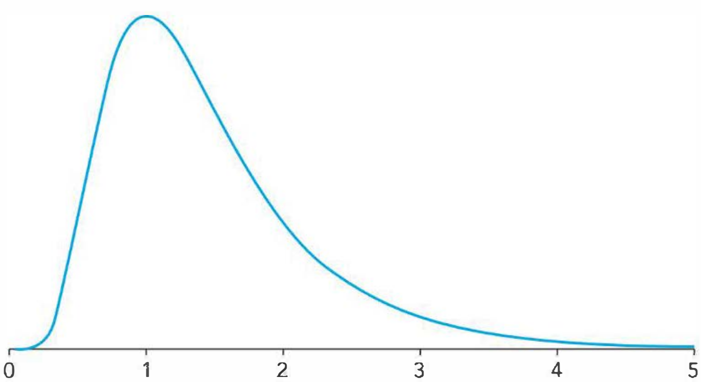


Figure 15.2 Lognormal distribution.

The stock we considered earlier had a price of USD 100, a mean return of 12% per year, and a volatility of 25%. As an approximation, we can reasonably assume the stock price at the end of a short period (e.g., one week) is normal.[4] When longer periods are considered, however, we must use Equations (15.2) and (15.3).

Suppose we are interested in a confidence interval for the stock price at the end of two years. From Equation (15.2), the mean of the logarithm of the stock price at the end of two years is

$$ln(100) + \left(0.12 - \frac{0.25^2}{2}\right) \times 2 = 4.783$$

Meanwhile, the standard deviation of the logarithm of the stock price at the end of two years is

$$0.25 \times \sqrt{2} = 0.354$$

[4] Recall that this is what we did when deriving confidence levels in earlier chapters.

Because the logarithm of the stock price is normally distributed, we know the 95% confidence interval for the logarithm of the stock price is

$$4.783 - 1.96 \times 0.354 \leq ln(S_T) \leq 4.783 + 1.96 \times 0.354$$

or

$$4.090 \leq ln(S_T) \leq 5.476$$

This means that:

$$e^{4.090} \leq S_T \leq e^{5.476}$$

or

$$59.7 \leq S_T \leq 238.8$$

Return Calculations

The realized return R from the stock in time T (when measured with continuous compounding) is given by:

$$S_T = S_0 e^{RT}$$

so that:

$$R = \frac{1}{T} ln\left(\frac{S_T}{S_0}\right)$$

This is normally distributed with mean:

$$\mu - \frac{\sigma^2}{2}$$

and standard deviation:

$$\frac{\sigma}{\sqrt{T}} \quad (15.5)$$

In the model, we are assuming the expected return over a very short period (theoretically, an infinitesimally short period) is μ.

However, the results we have just presented show the return over a finite time period of length T is $\mu - \frac{\sigma^2}{2}$.

These results seem inconsistent. The reason they are, in fact, both correct is that the return in time T is not the arithmetic average of the returns over n short periods of length Δt, when $n = T/\Delta t$. To illustrate this, suppose USD 100 is invested for three months and the returns per month (with monthly compounding) are 4%, 10%, and −8%. The value of the investment at the end of three months is

$$100 \times 1.04 \times 1.1 \times 0.92 = 105.25$$

This is $100 \times (1.0172)^3$, showing that the investor has earned 1.72% per month with monthly compounding. This is less than the average monthly return of 2% (= (4% + 10% − 8%)/3).[5]

Imagine making the length of each period (one month in our example) progressively smaller. In the limit, we find the expected return over time T with continuous compounding is less than the expected return over an infinitesimally short period. The former is $\mu - \sigma^2/2$, whereas the latter is μ.[6]

15.2 VOLATILITY

Volatility (denoted by σ) is a measure of our uncertainty about the returns provided by an investment. From Equation (15.5), it can be defined as the annualized standard deviation of the return (measured with continuous compounding).

We discussed the use of historical data to estimate volatility in Chapter 3. We noted that risk managers often approximate volatility per day as the square root of the average squared daily return. This approximation works reasonably well for daily data. Here we describe how a more precise estimate can be obtained.[7]

[5] This result can be related to the difference between geometric and arithmetic averages. The geometric average of n numbers is the nth root of the product of the numbers. For example, the geometric average of 2, 3, and 4.5 is 3, whereas, the arithmetic average is 3.1667. The geometric average of a set of numbers (not all equal) is always less than the arithmetic average. If R is the return over n periods, and r_i is the return in the ith period, $1 + R$ is the geometric average of $1 + r_i$. If R were an arithmetic average of the returns, $1 + R$ would be the arithmetic average of $1 + r_i$.

[6] Equation (15.4) is related to this result. It implies that $\ln(E(S_T/S_0)) > E(\ln(S_T/S_0))$. From Equation (15.1), the left hand side of this inequality equals μT while the right hand side is $E(\ln(S_{\Delta t}/S_0) + \ln(S_{2\Delta t}/S_{\Delta t}) + \cdots + \ln(S_T/S_{T-\Delta t}))$, which is in the limit as Δt tends to zero becomes $(\mu - \sigma^2/2)T$.

[7] The more accurate approach should be used when the time between observations is longer than a day.

Suppose we collect data on the price of an asset (e.g., a stock) at intervals of τ years (e.g., for monthly data, $\tau = 1/12$; for weekly data, $\tau = 1/52$). Define S_i as the ith observation ($0 \le i \le n$) and:

$$u_i = ln\left(\frac{S_i}{S_{i-1}}\right) \quad i = 1, 2, \ldots, n$$

Let s be the standard deviation of u_i. From Equation (15.5), an estimate of the volatility per year is

$$\frac{s}{\sqrt{\Delta t}}$$

The standard error of this estimate can be shown to be approximately equal to the estimate divided by the square root of $2n$.

As an illustration of these calculations, consider the weekly observations on a stock price in the second column of Table 15.1. The price relatives (i.e., the price in the current week divided by the price in the previous week) are in the third column. The logarithm of the price relative is shown in the fourth column.

In this case:

$$\sum_{i=1}^{10} u_i = 0.1398$$

$$\sum_{i=1}^{10} u_i^2 = 0.0145$$

The average value of u_i ($\bar{u}$) is 0.01398 (= 0.1398/10) and the usual formula for standard deviation gives the standard deviation of u_i as:

$$\sqrt{\frac{1}{9}\left(\sum_{i=1}^{10}(u_i^2 - \bar{u}^2)\right)} = \sqrt{\frac{1}{9}(0.0145 - 10 \times 0.01398^2)} = 0.0373$$

Table 15.1 Volatility Calculation

Week, i	Stock Price, S_i	S_i/S_{i-1}	$u_i = \ln(S_i/S_{i-1})$
0	40.0		
1	41.0	1.0250	0.0247
2	43.0	1.0488	0.0476
3	41.5	0.9651	−0.0355
4	39.0	0.9398	−0.0621
5	41.0	1.0513	0.0500
6	42.5	1.0366	0.0359
7	42.5	1.0000	0.0000
8	43.0	1.0118	0.0117
9	45.0	1.0465	0.0455
10	46.0	1.0222	0.0220

The volatility per year is estimated as:

$$\frac{0.0373}{\sqrt{1/52}} = 0.269$$

or 26.9%. The standard error of this estimate is about 6.02% (= 26.9%/$\sqrt{(2 \times 10)}$).

Dividends must be considered when the volatility of an individual stock is calculated from historical data. When a dividend is declared on a stock, an ex-dividend date is specified. Investors who own the stock before the ex-dividend date receive the dividend, whereas those who own the stock after the ex-dividend date do not. The stock price therefore declines on the ex-dividend date. Tax considerations also play a part in the size of the decline. The safest approach is to remove stock price changes on ex-dividend dates from the sample data used to estimate volatility.

Measuring Time

We might expect the variance of the return provided by an asset over a three-day period to be three multiplied by the variance over a one-day period. However, research has shown this is not the case when the three-day period includes a two-day weekend.[8] The variance of the return calculated between the close of trading Friday and close of trading Monday is much less than three times that observed during a weekday when markets are open.

While this phenomenon may change as trading outside normal trading hours becomes more common, the current convention in derivatives markets is to assume volatility is a trading time phenomenon rather than a calendar time phenomenon. When volatility is calculated and options are valued, time is therefore measured in trading days (i.e., days when the market is open). There are usually assumed to be 252 trading days in a year so that:[9]

$$\text{Volatility per Year} = \text{Volatility per Day} \times \sqrt{252}$$

The life of an option is calculated in trading days. For example, if there are 56 trading days to maturity, the life is calculated as 0.222 (= 56/252) years.

[8] See E. F. Fama, "The Behavior of Stock Market Prices," Journal of Business, 38 (January 1965): 34–105; K. R. French, "Stock Returns and the Weekend Effect," Journal of Financial Economics, 8 (March 1980): 55–69; K. R. French and R. Roll, "Stock Return Variances: The Arrival of Information and the Reaction of Traders" *Journal of Financial Economics*, 17 (September 1986): 5–26; R. Roll "Orange Juice and the Weather," American Economic Review, 74, 5 (December 1984): 861–880.

[9] Currencies tend to be traded for more days per year than other assets. As a result, it is usually assumed that there are 262 trading days per year when currency options are valued.

15.3 NO-ARBITRAGE FOR OPTIONS

The assumptions necessary to derive the Black-Scholes-Merton options pricing model are as follows.

- The behavior of stock prices corresponds to the model in Section 15.1 with μ and σ held constant.
- There are no transaction costs or taxes and all securities are perfectly divisible.
- There are no dividends on the stock during the life of the option.
- There are no riskless arbitrage opportunities.
- Security trading is continuous.
- Investors can borrow or lend at the same risk-free rate, which is constant through time.
- The options being considered cannot be exercised early.

Some of these assumptions have been relaxed in subsequent research. For example, r and σ can be functions of time, and (as we will see later in this chapter) some results can be produced for situations where dividends are anticipated.

In this section, we focus on the nature of the no-arbitrage argument. This is at the heart of Merton's derivation and is important for the way derivatives are hedged. Define the price of a call option as c and the underlying stock price as S. As in Chapter 14, let Δ be the sensitivity of the call option price to the stock price so that:

$$\Delta = \frac{\delta c}{\delta S}$$

where δS is a small change in the stock price, and δc is the corresponding change in the call option price. The delta of a call option is the gradient illustrated by the dotted line in Figure 15.3. We do not know either c or Δ, but we do know that we have a riskless position if we sell one call option and buy Δ shares of the stock. The cost of setting up the position is

$$S\Delta - c$$

The position should earn the risk-free rate over the next short period of time.

This is similar to the argument we used in Chapter 14 when we were valuing options using binomial trees. In Chapter 14, the option position remained riskless for a single step of the tree. In this case, the period of time over which the option remains riskless is (in theory) infinitesimally short. This is because c is a continuous curve as a function of S (as indicated in Figure 15.3). However, Merton was able to derive a differential equation that must be satisfied by c.

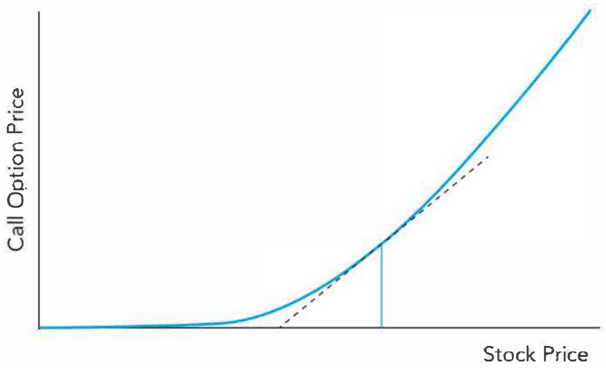


Figure 15.3 The solid line shows the call option price. The gradient of the dotted line is the Δ of the option for the stock price indicated by the vertical line.

So far, we have not said anything about the call option's payoff or its time to maturity. To use Merton's result to value a European option, we must apply what are termed boundary conditions. The key boundary condition for a European call option with time to maturity T and strike price K is that the value of the option is $\max(S - K, 0)$ at time T. For a European put, this boundary condition is $\max(K - S, 0)$. Other derivatives give rise to other boundary conditions (some quite complicated).

15.4 THE PRICING FORMULAS

The Black-Scholes-Merton formulas for the price of a European call option (c) and a European put option (p) are

$$c = S_0 N(d_1) - Ke^{-rT} N(d_2) \tag{15.6}$$

$$p = Ke^{-rT} N(-d_2) - S_0 N(-d_1) \tag{15.7}$$

Where:

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

$$d_2 = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}$$

In these formulas, S_0 is the current stock price, K is the strike price, T is the time to maturity in years, r is the risk-free rate per year (continuously compounded) for a maturity of T, and σ is an estimated volatility per year over the next T years. The N function is the cumulative normal distribution function (which we have used in earlier chapters), and it can be calculated from NORMSDIST in Excel or from tables.

As an example of the formula, suppose the stock price is USD 56, the strike price is USD 60, the risk-free rate is 5% per annum, the volatility is 30% per annum, and the time to maturity is 18 months. In this case:

$$d_1 = \frac{\ln(56/60) + (0.05 + 0.3^2/2) \times 1.5}{0.3\sqrt{1.5}} = 0.2001$$

$$d_2 = \frac{\ln(56/60) + (0.05 - 0.3^2/2) \times 1.5}{0.3\sqrt{1.5}} = -0.1674$$

The price of the European call option is

$$56N(0.2001) - 60e^{-0.05\times 1.5} \times N(-0.1674) = 8.3069$$

The price of a European put option with the same strike price is

$$60e^{-0.05\times 1.5} \times N(0.1674) - 56N(-0.2001) = 7.9715$$

We introduced put-call parity when discussing the properties of stock options in Financial Markets and Products. This states that when a European put option and a European call option on a non-dividend paying stock have the same strike price and time to maturity:

$$\text{Call Price} + \text{PV of Strike Price} = \text{Put Price} + \text{Stock Price}$$

Put-call parity is satisfied in our example because:

$$8.3069 + 60e^{-0.05\times 1.5} = 63.9715 = 56 + 7.9715$$

15.5 RISK-NEUTRAL VALUATION

We introduced risk-neutral valuation in connection with binomial trees in Chapter 14. It is a general principle stating that we can price derivatives on the assumption all investors are risk neutral (i.e., they do not adjust their required expected returns for risk). We get the correct price for not just a risk-neutral world, but for the real world as well.

Using a risk-neutral approach to price a derivative dependent on a traded security requires that we:

- Assume that the expected return from the underlying asset is the risk-free rate,
- Calculate the expected payoff from the derivative, and
- Discount the expected payoff at the risk-free rate.

Consider a derivative that provides a payoff V at time T, where V is a function of market variables (e.g., equity prices and exchange rates). The price of the derivative is

$$e^{-rT}\hat{E}(V)$$

where $\hat{E}$ denotes expected payoff in a risk-neutral world.

The Black-Scholes-Merton price for a European call option can be deduced by evaluating:

$$e^{-rT}\hat{E}(\max(S_T - K, 0))$$

The price of a European put option can be deduced by evaluating:

$$e^{-rT}\hat{E}(\max(K - S_T, 0))$$

Evaluating the expectations in these expressions requires some messy calculus. We will therefore illustrate risk-neutral valuation with a simpler example. Consider a forward contract to buy a non-dividend paying stock for K at time T. The payoff is $S_T - K$ and the value of the instrument is

$$f = e^{-rT}\hat{E}(S_T - K)$$

Because K is a constant, this becomes

$$f = e^{-rT}\hat{E}(S_T) - Ke^{-rT}$$

In a risk-neutral world, the expected stock price at time T is given by setting $\mu = r$ in Equation (15.1). Hence:

$$\hat{E}(S_T) = S_0e^{rT}$$

and

$$f = S_0 - Ke^{-rT}$$

This is the formula for the value of a forward contract on an investment asset that provides no income, which was deduced (using a different approach) in Financial Markets and Products.

15.6 IMPLIED VOLATILITY

The implied volatility of an option is the volatility that gives the market price of the option when it is substituted into the Black-Scholes-Merton formula. There is no analytic formula for implied volatility. Instead, it must be found using an iterative trial and error procedure.

One simple approach to determining implied volatility is called successive bisection. Note that the price of an option is a continuous increasing function of its volatility. This means that if we keep increasing volatility, we will find a volatility that is higher than the implied volatility (i.e., a volatility that gives a value for the option that is greater than the market price). We will refer to this as the "too high volatility." Similarly, a volatility of zero is lower than the implied volatility since it gives a price that is less than the market price.[10] We will refer to this as the "too low volatility." We then try a volatility that is the average of the too high and too low volatilities. If it gives a price that is too high, this volatility becomes the new too high volatility. If it gives a price that is too low, it becomes the new too low volatility. The new too high and too low volatilities are then averaged, and the procedure is repeated until the implied volatility is found.[11]

Consider again the option in Section 15.4. The stock price is USD 56, the strike price is USD 60, the risk-free rate is 5%, and the time to maturity is 1.5 years. Suppose the market price of the option is USD 7.00. The implied volatility is the value of σ that gives a price of USD 7.00 when substituted into Equation (15.6).[12]

We know from Section 15.4 that a volatility of 30% is too high because it gives a price of USD 8.3069 (which is greater than the market price of USD 7.00). If we change the volatility to 20%, the price becomes USD 5.6117 (which is too low). We therefore know that the implied volatility is somewhere between 20% and 30%. Next, we try 25%. This gives an option price of USD 6.9624 (which is again too low). The implied volatility therefore lies between 25% and 30%. We next try 27.5%. This gives USD 7.6355 (which is again too high). We now know that the implied volatility lies between 25% and 27.5%. Continuing in this way, we find that the implied volatility is 25.14%. When this volatility is substituted into Equation (15.6), we get a price of USD 7.00.

Implied volatilities can be similarly calculated for American options valued using the binomial tree approach shown in Chapter 14. As mentioned in an earlier chapter, the Chicago Board Options Exchange has developed indices that track volatilities. The most popular of these is the SPX VIX index, which tracks the volatilities of 30-day options on the S&P 500. Other indices track the volatilities of commodities, interest rates, currencies, and other stock indices. There is even a volatility of volatility index, which tracks the volatility of the VIX itself. (This is called the VVIX index.)

Traders monitor implied volatilities carefully and often use them to communicate prices. If the assumptions underlying the Black-Scholes-Merton model held exactly, all options on an asset would have the same implied volatility at all times. In practice, implied volatilities vary with strike prices. This indicates that the market does not price options consistently with the Black-Scholes-Merton assumptions. Figures 15.4 and 15.5 show patterns that are typically observed for the implied volatilities of options on currencies and equities with a particular maturity. The patterns are referred to as *volatility smiles.*[13]

[10] If this is not the case, a trader has an arbitrage opportunity. Lower bounds for options and associated arbitrage opportunities were discussed when we considered the properties of stock options in Financial Markets and Products.

[11] This is a straightforward procedure. Other procedures that are more numerically efficient for solving a non-linear equation are used in practice.

[12] RMFI software that accompanies John Hull's book "Risk Management and Financial Institutions" fifth edition, 2018, Wiley can be used to implement the valuation formulas in this chapter and the calculation of implied volatilities. See www-2.rotman.utoronto.ca/~hull/riskman

[13] In the case of equities, the volatility smile is not symmetrical and is sometimes referred to as a volatility skew.

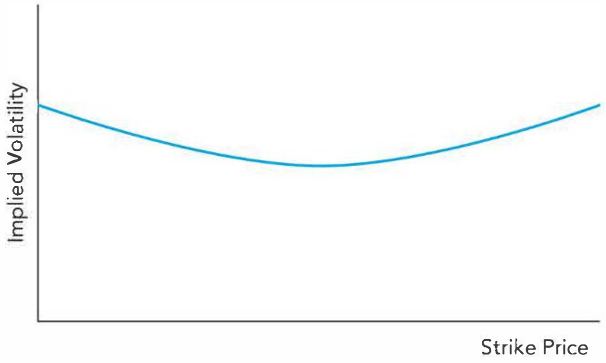


Figure 15.4 Typical volatility smile for options on a foreign currency. All options have the same maturity.

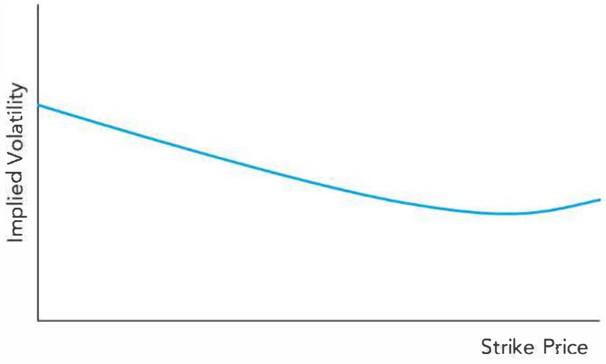


Figure 15.5 Typical volatility smile for an equity option. All options have the same maturity.

More generally, traders monitor the *volatility surface*. This describes implied volatilities as a function of both strike price and time to maturity. When quoting options prices, traders interpolate between the known implied volatilities to determine an implied volatility for the option under consideration. This is then substituted into the Black-Scholes-Merton equation to determine the option price. This procedure is a way of overcoming the fact that the market does not price options consistently with the Black-Scholes-Merton assumptions.

15.7 OPTIONS ON STOCK INDICES, CURRENCIES, AND FUTURES

In Chapter 14, we first described how the binomial tree methodology can be used for non-dividend paying stocks before examining how it could be extended to value options on stock indices, currencies, and futures. So far in this chapter, we have only considered options on non-dividend paying stocks. However, if we assume other assets follow a similar process to that assumed for stock prices in Section 15.1, the Black-Scholes-Merton results can also be extended to those assets. In all cases, volatilities can be calculated from historical data in the same way as for non-dividend paying stocks. Implied volatilities are also defined in the same way as for non-dividend paying stocks.[14]

For a European option on a stock paying a continuous dividend yield at rate q, Equations (15.6) and (15.7) become

$$c = S_0e^{-qT}N(d_1) - Ke^{-rT}N(d_2) \tag{15.8}$$

$$p = Ke^{-rT}N(-d_2) - S_0e^{-qT}N(-d_1) \tag{15.9}$$

where

$$d_1 = \frac{\ln(S_0/K) + (r - q + \sigma^2/2)T}{\sigma\sqrt{T}}$$

$$d_2 = \frac{\ln(S_0/K) + (r - q - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}$$

These formulas can be used for a European option on a stock index paying dividends at rate q when S_0 is the value of the index.

When considering an option on a foreign currency, we recognize that it behaves like a stock paying a dividend yield at the foreign risk-free rate (r_f). The valuation equations are therefore obtained by setting $q = r_f$ in Equations (15.8) and (15.9). They are

$$c = S_0e^{-r_fT}N(d_1) - Ke^{-rT}N(d_2) \tag{15.10}$$

$$p = Ke^{-rT}N(-d_2) - S_0e^{-r_fT}N(-d_1) \tag{15.11}$$

where S_0 is the current exchange rate, σ is its volatility, and:

$$d_1 = \frac{\ln(S_0/K) + (r - r_f + \sigma^2/2)T}{\sigma\sqrt{T}}$$

$$d_2 = \frac{\ln(S_0/K) + (r - r_f - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}$$

When considering an option on futures, we recognize that a futures price (F) behaves like a stock paying a dividend yield at the domestic risk-free rate (r). The valuation equations are therefore obtained by setting $q = r$ and $S_0 = F_0$ in Equations (15.8) and (15.9). They are

$$c = F_0e^{-rT}N(d_1) - Ke^{-rT}N(d_2) \tag{15.12}$$

$$p = Ke^{-rT}N(-d_2) - F_0e^{-rT}N(-d_1) \tag{15.13}$$

where:

$$d_1 = \frac{\ln(F_0/K) + \sigma^2T/2}{\sigma\sqrt{T}}$$

$$d_2 = \frac{\ln(F_0/K) - \sigma^2T/2}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}$$

[14] There are many other derivatives whose prices can be calculated analytically when the assumptions underlying the Black-Scholes-Merton model are made. Examples are barrier options, lookback options, and perpetual American options. See J. Hull, "Options, Futures, and Other Derivatives," 10th edition, Pearson, 2018, Chapter 26.

Equations (15.12) and (15.13) are referred to as Black's model.[15] The volatility parameter here (σ) is the volatility of the futures price.

As discussed in Financial Markets and Products, the forward price for a forward contract and the futures price for the corresponding futures contract can be assumed to be approximately the same in many situations. Black's model therefore also gives the price of an option on a forward contract. Furthermore, a European put (call) option on the spot price of an asset is the same as a European put (call) option on the forward price of the asset when:

- They have the same strike price and time to maturity, and
- The forward contract matures at the same time as the option.

This is because the forward price equals the spot price at the maturity of the forward contract.

This result allows Black's model to be used to value an option on the spot price of an asset in terms of the forward price of the asset. This is an approach often taken by traders because it avoids the need to estimate income on the asset (since the forward price captures all relevant aspects of the income).

15.8 HANDLING DISCRETE DIVIDENDS

Up to now, we have assumed that the stock underlying the option pays no dividends. We now relax this assumption and assume the dividends paid during the life of the option are either known for certain or can be estimated with reasonable accuracy. This is not an unreasonable assumption for options lasting for a short period of time (i.e., less than one or two years). For longer maturity options, it is often assumed the dividend yield on the stock is known. The option can then be valued using Equations (15.8) and (15.9).

Before proceeding further, a point should be made about the measurement of dividends. The relevant dividends for valuing an option are those for which the ex-dividend date is during the life of the option. The relevant size of each dividend is the amount by which the stock price is expected to decrease on the ex-dividend date due to the dividend. The latter may be different from the amount of the dividend declared if capital gains are taxed differently from investment income.[16]

[15] See F. Black, "The Pricing of Commodity Contracts," *Journal of Financial Economics*, 3 (1976): 167–179.

[16] The effect of a dividend is to provide the owner of the stock with investment income while the reduction in the stock price leads to a capital loss.

Suppose the present value of the dividends paid during the life of a European option is estimated as D. The current stock price (S) can be assumed to have two components:

1. A component that will be used to pay the dividends (D), and
2. A component that will still exist at option maturity ($S - D$).

If we assume that the volatility σ applies to the second component, the option can be valued by replacing S with $S - D$ in Equations (15.6) and (15.7).

As an illustration, consider a six-month European put option on a stock whose price is USD 74. The strike price is USD 70, the risk-free interest rate is 5% per year (continuously compounded), and the volatility parameter is 20%. Additionally, dividends of USD 1.50 are expected with ex-dividend dates in one month and four months.

In this case, the present value of the dividends is

$$1.50e^{-0.05\times0.08333} + 1.5e^{-0.05\times0.33333} = 2.9690$$

The option can therefore be valued using Equation (15.7) with:

$$S = 74 - 2.9690 = 71.0310$$

The variables d_1 and d_2 are

$$d_1 = \frac{ln(71.0310/70) + (0.05 + 0.2^2/2) \times 0.5}{0.2\sqrt{0.5}} = 0.3509$$

$$d_2 = d_1 - 0.2\sqrt{0.5} = 0.2095$$

The option price is therefore:

$$70e^{-0.05\times0.5}N(-0.2095) - 71.0310N(-0.3509) = 2.70$$

American Options

So far, all the options considered in this chapter have been European. As we saw in Financial Markets and Products, call options on a non-dividend paying stock should never be exercised early. It follows that Equation (15.6) provides prices for American call options on non-dividend paying stocks as well as for European call options.

It may be optimal to exercise early when there are discrete dividends, but only immediately before an ex-dividend date. Suppose the ex-dividend dates during the life of an option are at times t_i ($1 \le n$) and the option matures at time T. Define D_i as the ith dividend. It can be shown that it is never optimal to exercise at time t_i ($1 \le i \le n - 1$) if:[17]

$$D_i \le K[1 - e^{-r(t_{i+1} - t_i)}]$$

[17] In the case of a small continuous dividend yield, it is, in theory, always optimal to exercise early for a sufficiently high value of the stock price. This is because the dividend paid during a small time interval Δt is $Sq\Delta t$, and for a sufficiently high S, it is worth exercising to obtain the stock and receive this dividend.

It is never optimal to exercise at time t_n if:

$$D_n \leq K[1 - e^{-r(T - t_n)}]$$

If dividends are sufficiently small so that these conditions are satisfied, an American call option on a dividend-paying stock can be valued as a European call option.

American put options on stocks and all American options on stock indices, currencies, and futures should not be valued as European options. Binomial trees can be used for valuation in these cases.

15.9 WARRANTS

Warrants are options issued by a company on its own stock. If warrants are exercised, the company issues more shares, and the warrant holder buys the shares from the company at the strike price. An option traded by an exchange does not change the number of shares issued by a company. However, a warrant allows new shares to be purchased at a price lower than the current market price. This dilutes the value of existing shares.

Assuming markets are efficient, the share price will reflect the potential dilution from outstanding warrants, which therefore does not have to be considered when those outstanding warrants are valued. This means that warrants can be valued in the same way as exchange-traded options once they have been announced.

A firm that is deciding whether or not to issue warrants might be interested in calculating the cost of the warrants to its existing shareholders. If there are N existing shares, and the company is contemplating the issue of M warrants (with each giving the warrant holder the right to buy one new share), it can be shown that the cost of each warrant to existing shareholders is $N/(N + M)$ multiplied by the price of each warrant. The price of each warrant is calculated in the same way as the price of an exchange-traded option and is based on the share price before the warrant issue was announced.

SUMMARY

A common assumption in the pricing of stock options is that the return on a stock over a very short period of time is normally distributed. This leads to the stock price at some future time having a lognormal distribution (i.e., the logarithm of the stock price is normal).

A key parameter necessary to estimate the price of a stock option is volatility. This can be defined as the standard deviation of the continuously compounded return on the stock during a one-year period. Volatility can be estimated from historical data. Traders often imply volatilities from options prices observed in the market. Typically, not all options have the same implied volatilities. This leads to, what are termed, volatility smiles and volatility surfaces.

The procedures that can be used to derive the Black-Scholes-Merton option pricing formulas parallel those used in connection with binomial trees discussed in Chapter 14. An instantaneously riskless position can be built using an option and the underlying stock. This means that it instantaneously earns the risk-free rate and leads to a differential equation, which the option price must satisfy. Risk-neutral valuation, which was introduced in Chapter 14, can also be used for option valuation.

The formulas for valuing European options on stock indices, currencies, and futures are small variations on the Black-Scholes-Merton formula. American options must usually be valued using binomial trees. Exceptions are American call options on non-dividend paying stocks, or stocks where dividends are low. These should never be exercised early and are therefore the same as their European counterparts.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

QUESTIONS

Short Concept Questions

15.1 What does the Black-Scholes-Merton model assume about stock price movements?

15.2 What is the relationship between the lognormal distribution and the normal distribution?

15.3 Provide a precise definition of the volatility of the price of a non-dividend paying stock.

15.4 What is the formula for the expected stock price at a future time when the Black-Scholes-Merton assumptions are made?

15.5 What is the formula for the expected return (with continuous compounding) over a time period of T when the Black-Scholes-Merton assumptions are made?

15.6 If you have five years of monthly data on a variable, how would you calculate its volatility?

15.7 Why do option traders assume 252 days per year?

15.8 What are the steps in valuing a derivative that provides a payoff at time T using risk-neutral valuation?

15.9 What is meant by a volatility surface?

15.10 How is the Black-Scholes-Merton formula used to value European options on a dividend-paying stock?

Practice Questions

15.11 A stock has an expected return of 15% and a volatility of 20%. The current price of the stock is USD 50. Estimate a 99% confidence interval for the price at the end of one day.

15.12 For the stock in Question 15.11, estimate 99% confidence for the price in six months.

15.13 For the stock in Question 15.11, what is the mean and standard deviation of the continuously compounded return over three years?

15.14 Monthly stock prices are as follows in USD: 35, 38, 41, 37, 33, and 32. Use this data to estimate the volatility per year.

15.15 If the volatility per day is 2%, what is the volatility per year?

15.16 A stock price is USD 50 with a volatility of 22%. The risk-free rate is 3%. Use the Black-Scholes-Merton formula to value (a) a European call option and (b) a European put option when the strike price is USD 50, and the time to maturity is nine months.

15.17 Use risk-neutral valuation to value a derivative that pays off $\ln(S_T)$ at time T, where S_T is the price of a non-dividend paying stock at time T. Define other variables as in this chapter.

15.18 The current exchange rate for a currency is 1.2000 and the volatility of the exchange rate is 10%. Calculate the value of a call option to buy 100 units of the currency in two years at an exchange rate of 1.2500. The domestic and foreign risk-free interest rates are 3% and 5%, respectively.

15.19 The futures price of an asset is USD 20, and the volatility of the futures price is 30%. Calculate the value of a put option to sell futures in three months for USD 22. The risk-free rate is 4%.

15.20 A stock provides dividends of USD 0.25 per share every three months. The next two ex-dividend dates are in two months and five months. The risk-free rate is 4% per annum. At what times might a six-month American call option with a strike price of USD 50 be exercised?

ANSWERS

Short Concept Questions

15.1 The return over any very short period of length Δt is normally distributed with mean $\mu\Delta t$ and standard deviation $\sigma\sqrt{\Delta t}$ where μ is the expected return and σ is the volatility. In theory, this is true only in the limit as Δt tends to zero.

15.2 A variable X has a lognormal distribution if $\ln(X)$ has a normal distribution.

15.3 Volatility is the standard deviation of the continuously compounded return over one year.

15.4 The expected stock price at time T is $S_0e^{\mu T}$ where S_0 is the stock price today, and μ is the expected return.

15.5 The expected return is $\mu - \sigma^2/2$. (This does not depend on T.)

15.6 If S_i is the value of the variable at the end of month i, the volatility is $\sqrt{12}$ times the standard deviation of the 59 values of $\ln(S_i/S_{i-1})$.

15.7 Volatility has historically been much lower when markets are closed than when they are open. There are about 252 days each year when markets are open.

15.8 We calculate the expected payoff from the derivative on the assumption that the return from the underlying asset is the risk-free rate. This expected payoff is then discounted at the risk-free rate.

15.9 The volatility surface shows the implied volatility of an option on an asset as a function of its strike price and time to maturity.

15.10 The present value of the dividends that have ex-dividend dates during the life of the option is subtracted from the stock price when the formula is used. The volatility applies to the stock price minus the present value of the dividends. In this context, a dividend is defined as the reduction in the stock price on the ex-dividend date.

Solved Problems

15.11 Here, we are dealing with a short time period, and so it is reasonable to assume that the return is normally distributed. The return has a mean of $15\% \times (1/252) = 0.0595\%$, and a standard deviation of $20\% \times \sqrt{1/252} = 1.2599\%$. The 99% confidence interval for the percentage return is between:

$$0.0595 - 1.2599 \times N^{-1}(0.995) = -3.186\%$$

and

$$0.0595 + 1.2599 \times N^{-1}(0.995) = +3.305\%$$

The confidence interval for the stock price is therefore between $50 \times 0.96814 = 48.4$ and $50 \times 1.03305 = 51.7$.

15.12 Here, the time period is longer, and we should work with lognormal distributions. From Equations (15.2) and (15.3), the logarithm of the stock price has mean:

$$\ln(50) + (0.15 - 0.20^2/2) \times 0.5 = 3.9770$$

and standard deviation:

$$0.2\sqrt{0.5} = 0.1414$$

We are 99% certain that:

$$3.9770 - N^{-1}(0.995) \times 0.1414 < \ln(S_T) < 3.9770 + N^{-1}(0.995) \times 0.1414$$

or

$$3.6127 < \ln(S_T) < 4.3413$$

so that:

$$e^{3.6127} < S_T < e^{4.3413}$$

or

$$37.1 < S_T < 76.8$$

15.13 The mean return (annualized) is $0.15 - 0.2^2/2 = 0.13$ or 13%. From Equation (15.5) the standard deviation of the return is $0.2/\sqrt{3} = 0.1155$ or 11.55%.

15.14 The calculations are in the following table. In this case, the average value of the u_i is -0.01792 and $\sum_{i=1}^{5} u_i^2 = 0.03711$ and the standard deviation of the u_i is

$$\sqrt{\frac{1}{4}(0.03711 - 5 \times 0.01792^2)} = 0.0942$$

so that the volatility is $0.0942\sqrt{12} = 0.326$ or 32.6%.

Month	Stock Price, S_i	S_i/S_{i-1}	$u_i = \ln(S_i/S_{i-1})$
0	35		
1	38	1.0857	0.0822
2	41	1.0789	0.0760
3	37	0.9024	−0.1027
4	33	0.8919	−0.1144
5	32	0.9697	−0.0308

15.15 The volatility per year is $2\% \times \sqrt{252} = 0.3175$ or 31.75%.

15.16 In this case:

$$d_1 = \frac{\ln(50/50) + (0.03 + 0.22^2/2) \times 0.75}{0.22 \times \sqrt{0.75}} = 0.2134$$

$$d_2 = \frac{\ln(50/50) + (0.03 - 0.22^2/2) \times 0.75}{0.22 \times \sqrt{0.75}} = 0.0228$$

and the call option price is

$$50N(0.2134) - 50e^{-0.03\times0.75}N(0.0228) = 4.3$$

The put option price is

$$50e^{-0.03\times0.75}N(-0.0228) - 50N(-0.2134) = 3.2$$

15.17 The value of the derivative is

$$e^{-rT}\hat{E}(\ln(S_T))$$

From Equation (15.2) this is

$$e^{-rT}\left[\ln(S_0) + \left(r - \frac{\sigma^2}{2}\right)T\right]$$

Note that we have replaced μ in Equation (15.2) with r.

15.18 In this case, $S_0 = 1.2000$, $K = 1.2500$, $r = 0.03$, $r_f = 0.05$, $\sigma = 0.1$, and $T = 2$, and Equation (15.10) gives

$$d_1 = \frac{\ln(1.2000/1.2500) + (0.03 - 0.05 + 0.1^2/2) \times 2}{0.1\sqrt{2}}$$
$$= -0.5008$$

$$d_2 = \frac{\ln(1.2000/1.2500) + (0.03 - 0.05 - 0.1^2/2) \times 2}{0.1\sqrt{2}}$$
$$= -0.6422$$

$$c = 1.2000e^{-0.05\times2}N(-0.5008) - 1.2500e^{-0.03\times2}N(-0.6422)$$
$$= 0.028$$

This is the value of an option to buy one unit of the currency. The value of an option to buy 100 units is 2.8.

15.19 In this case $F_0 = 20$, $K = 22$, $r = 0.04$, $\sigma = 0.3$, $T = 0.25$, and Equation (15.13) gives

$$d_1 = \frac{\ln(20/22) + (0.3^2/2) \times 0.25}{0.3\sqrt{0.25}} = -0.5604$$

$$d_2 = \frac{\ln(20/22) - (0.3^2/2) \times 0.25}{0.3\sqrt{0.25}} = -0.7104$$

$$p = 22e^{-0.04\times0.25}N(0.7104) - 20e^{-0.04\times0.25}N(0.5604) = 2.48$$

15.20 The possible exercise times are immediately before the stock goes ex-dividend at the two-month point and the five-month point, as well as at the end of the option's life. The ex-dividend dates are after 0.16667 and 0.41667 years, and the end of the option's life is after 0.5 years. Because:

$$0.25 < 50[1 - e^{-0.04\times(0.41667-0.16667)}]$$

we can use the results at the end of Section 15.7 to show that the option should never be exercised at the two-month point. Because:

$$0.25 > 50[1 - e^{-0.04\times(0.5-0.41667)}]$$

the option will sometimes be exercised immediately before the five-month point. Possible exercise times are therefore at five months and six months.

16

Option Sensitivity Measures: The "Greeks"

Learning Objectives

After completing this reading, you should be able to:

- Describe and assess the risks associated with naked and covered option positions.
- Describe the use of a stop loss hedging strategy, including its advantages and disadvantages, and explain how this strategy can generate naked and covered option positions.
- Calculate the delta of an option.
- Explain delta hedging for an option position, including its dynamic aspects.
- Define and describe vega, gamma, theta, and rho for option positions and calculate the gamma and vega of an option.
- Explain how to implement and maintain a delta-neutral and gamma-neutral position.
- Describe the relationship between delta, theta, gamma, and vega.
- Calculate the delta, gamma, and vega of a portfolio.
- Describe how to implement portfolio insurance and how this strategy compares with delta hedging.

Derivatives traders must manage the risks they take. The Greek letters (or Greeks, as they are sometimes called) are designed to provide information about these risks. Some Greek letters are concerned with movements in the price of the underlying asset; some are concerned with volatility changes; and others involve interest rates and dividend yields.

Typically, a trader is subject to limits on how large the Greek letters can be. If one of the Greek letters exceeds the applicable limit near the end of the trading day, a trader must either execute a trade that corrects the situation or seek special permission from the risk management function to maintain the existing position.

16.1 INTRODUCTION

Before describing the Greek letters, it will be instructive to consider several simple strategies for managing derivatives risk and explain why they are inadequate.

Suppose a trader has sold a client over-the-counter European call options to buy 1 million shares of a stock. The stock price is USD 100 per share, and the strike price is USD 105. The risk-free rate is 4% per annum, the volatility of the stock price is 25%, and the time to maturity is one year. The Black-Scholes-Merton pricing equation (15.6) shows that the theoretical price of an option to buy one share for USD 105 is USD 9.56.

Suppose the price obtained by the trader for 1 million call options is USD 10 million. The theoretical value of the transaction (USD) to the trader is

$$10{,}000{,}000 - 9{,}560{,}000 = 440{,}000$$

In this chapter, we consider how the trader should manage the risks in a position such as this. Three choices are

1. Buy 1 million options that are the same as those sold,
2. Do nothing, or
3. Cover the position by buying 1 million shares.

The first strategy would provide a perfect hedge. However, it is unlikely to be financially attractive. The trader would have to buy 1 million options with a strike price of USD 105 and a life of one year. The only market participant willing to take the other side of the transaction is likely to be a trader working for another bank.[1] This trader is likely to want to build a profit into his or her pricing, and it is therefore unlikely that the options could be purchased for much less than USD 10 million.

[1] The trader would like to find another end-user client who wants to take the opposite position (i.e., sell 1 million shares for USD 105 in one year) and would be prepared to do so for, say, USD 9 million. Although it does sometimes happen that traders working for banks find two clients who want to take opposite positions at the same time, this is a relatively rare occurrence.

The second strategy (sometimes referred to as a *naked position*) works well if the stock price stays below USD 105 for the next year. The options are then not exercised, and the trader's profit is USD 10 million. Given that the volatility of the stock price is 25%, however, it is quite possible it will rise to USD 125 or even USD 150 by the end of the year. The trader will then take a loss. A stock price of USD 125 will lead to a negative payoff of USD 20 million on the options and a net loss (after the amount received for the options is considered) of USD 10 million. A stock price of USD 150 at the end of the year would be even worse and lead to a net loss of USD 35 million.

The third strategy works well if the option is exercised. At the time the options are sold, the trader buys 1 million shares for USD 100 million. If the options are exercised, the trader receives USD 105 million for the shares. The total profit from selling the options and covering the position in this way is then USD 15 million. However, there is the potential for a considerable loss if the price declines. If it declines to USD 75 by the end of the year, the trader would lose USD 25 million on the purchased shares so that the net loss (after the price received for the option is considered) is USD 15 million. If the share price declines to USD 50 (which is possible), the net loss would be higher at USD 40 million.

Stop-Loss Strategy

Another possible hedging strategy is called the stop-loss strategy. This is a mixture of the second and third strategies we have just considered. Under this approach, the trader attempts to have a naked position when the option is out-of-the-money (i.e., the stock price is less than the strike price) and a fully covered position when it is in-the-money (i.e., the stock price is greater than the strike price).

The strike price in our example is USD 105. When the price of the stock is below USD 105, the trader maintains a naked position. As soon as the price rises above USD 105, the trader covers the position by buying 1 million shares. If the price subsequently falls below 105, the trader sells the shares. If it later rises above USD 105, the trader repurchases the shares.

If we assume the stock can be bought at USD 105 as it moves in-the-money and that it can be sold at USD 105 as it moves out-of-the-money, the stop-loss strategy leads to one of the following results.

- If at the end of the year the stock price is less than USD 105, the trader has a naked position and makes a USD 10 million profit. (Note that every time the stock is purchased at 105, it is later sold for 105.)

- If at the end of the year the stock price is more than USD 105, the trader has a fully covered position. The net result of all the trader's transactions is that he or she has bought 1 million shares for USD 105 million. When the options are exercised, the shares are sold for USD 105 million. Again, the trader makes a net profit of USD 10 million.

This strategy also appears to work well if the option is initially in-the-money. The trader must initially cover the position and then makes a profit corresponding to the difference between the price at which the option is sold and its initial intrinsic value.

Unfortunately for option traders, the stop-loss strategy is flawed. When the stock price reaches USD 105, the trader does not know whether the next price move will be up (so that the option moves in-the-money) or down (so that the option moves out-of-the-money). If the stock price moves from USD 104.9 to USD 105 and the trader buys 1 million shares, it is possible that the next price move will be back to USD 104.9. Similarly, if the stock price moves down from USD 105.1 to USD 105, and the trader sells 1 million shares, the next price move might be back up to 105.1.

As a practical matter, the stock must be bought when the price is $105 + e$ and sold when the price is $105 - e$ for some small value of e. If $e = 0.1$, for example, 1 million shares would be bought when the price is USD 105.1 and 1 million shares would be sold when the price is USD 104.9. This introduces a cost of USD 0.2 per share every time shares are bought and sold.

If the stock price never reaches the strike price of USD 105, the stop-loss strategy still works well. If it passes through the strike price level of USD 105 just once, the strategy also works well. However, if it passes through the strike price level many times, the strategy could be very expensive.

An obvious suggestion here is to make e smaller. In practice, the smallest possible value of e depends on how prices are quoted. But it can be shown that as e becomes smaller, the expected number of times that the stock is bought and sold increases.

Note that whatever the hedging strategy used the present value of the expected cost of hedging should always be the Black-Scholes-Merton value (USD 9.56 million in our example) provided the Black-Scholes-Merton assumptions hold. The hallmark of a good hedging strategy is that the cost of writing and hedging the options is always close to this expected cost. In our example, this would lead to a profit of close to USD 440,000 (= 10 million − 9.56 million) regardless where the stock price goes over the next year. The stop-loss strategy and other strategies we have considered so far are not good hedging strategies because the cost of writing and hedging the option is sometimes much more (or sometimes much less) than the theoretical price of the option.

16.2 DELTA HEDGING

The delta of an option (Δ) was introduced in Chapter 14 and mentioned again in Chapter 15. For the binomial trees considered in Chapter 14, buying Δ shares for each option sold led to a position that was riskless for a one-time step. In the continuous time model in Chapter 15, buying Δ shares for each option sold created a position that was riskless for a (theoretically instantaneously) short period of time.

The formula for the delta of a derivative dependent on a stock price S is[2]

$$\Delta = \frac{\delta f}{\delta S}$$

where δS is a small change in the stock price, and δf is the resultant change in the value of the derivative (with everything except the stock price being kept fixed).[3] Deltas for call and put options are illustrated in Figure 16.1.

The delta of a long position in a European call option on a non-dividend paying stock is

$$\Delta(\text{call}) = N(d_1) \tag{16.1}$$

The delta of a long position in a European put option on a non-dividend-paying stock is

$$\Delta(\text{put}) = N(d_1) - 1$$

The parameter d_1 is defined (as in the Black-Scholes-Merton equations) as:

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

where K is the strike price, T is the time to maturity, and r is the risk-free rate for maturity T. By convention, the volatility σ is set equal to the option's implied volatility (rather than a historical estimate of volatility) when all Greek letters are calculated. The function **N** is the cumulative normal distribution function for a standard normal variable with mean zero and standard deviation one.

The delta of a long position in a call option is positive since the price of a call option increases as the stock price increases. Figure 16.2 shows the variation of delta for a call option with respect to the variation in the stock price. When the stock price is very low relative to the strike price, the option is

[2] To be more precise we can use calculus notation and write $\Delta = \frac{\partial f}{\partial S}$.

[3] John Hull's RMFI software that accompanies his book, "Risk Management and Financial Institutions" fifth edition, 2018, Wiley, can be used to calculate all the Greek letters in this chapter. See www-2.rotman.utoronto.ca/~hull/riskman

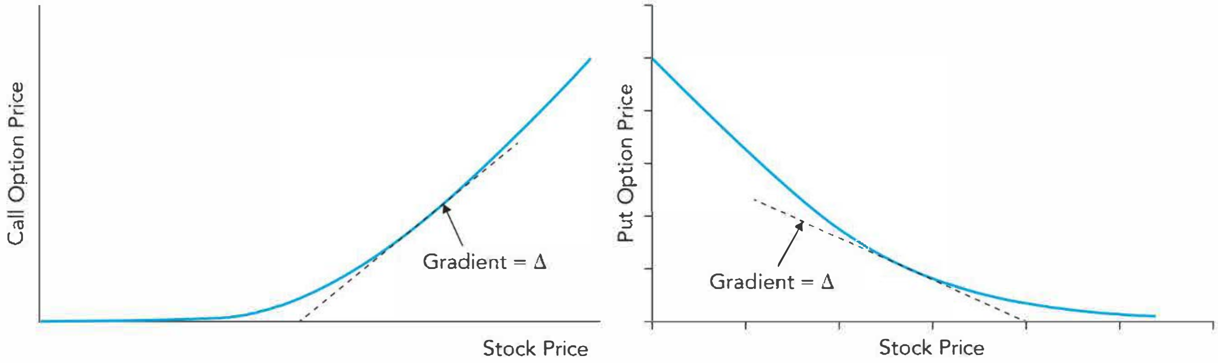


Figure 16.1 **Delta of call and put options.**

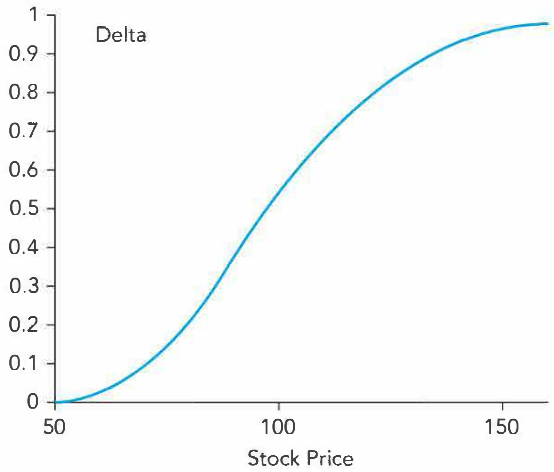


Figure 16.2 **Delta of a call option when the strike price is 105, the risk-free rate is 4%, the volatility is 25%, and the time to maturity is one year.**

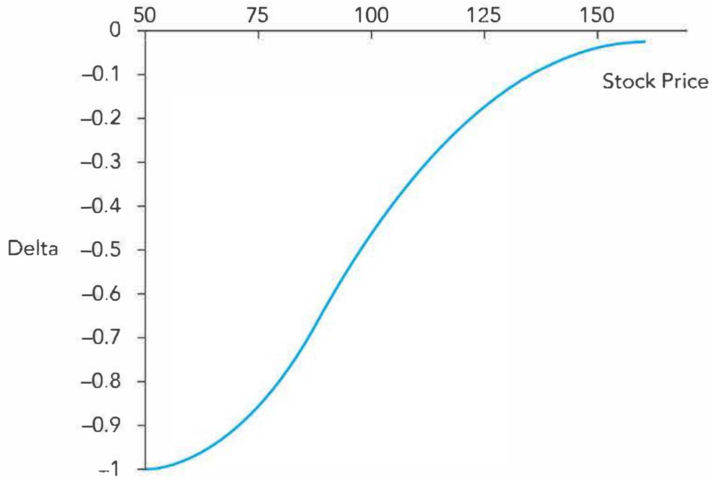


Figure 16.3 **Delta of a put option when the strike price is 105, the risk-free rate is 4%, the volatility is 25%, and the time to maturity is one year.**

deep-out-of-the-money and is worth close to zero. Thus, a small change in the stock price has very little effect on the option price (since it remains deep-out-of-the-money). As a result, the delta is close to zero in Figure 16.2 for low stock prices.

Now consider the situation where the stock price is very high relative to the strike price. In this case, the option is deep-in-the-money and is almost certain to be exercised. The option holder's position can be thought of as being almost the same as one where he or she owns stock and will pay K at time T. When the stock price increases (or decreases) by δS, the option price therefore increases (or decreases) by close to δS. This explains why the delta is close to one for high stock prices in Figure 16.2.

The delta of a long position in a put option is negative because the price of a put option decreases as the stock price increases. Figures 16.3 shows the variation of delta for a put option with respect to the variation in the stock price. In the case when the stock price is very low relative to the strike price, the option will almost certainly be exercised. When the stock price increases (or decreases) by δS, the option price decreases (or increases) by close to δS. This explains why the delta is close to -1 for low stock prices in Figure 16.3. When the stock price is very high relative to the strike price, a put option is deep-out-of-the-money. It continues to be deep-out-of-the-money when the stock price changes by a small amount. As a result, delta is close to zero.

Delta Hedging

Consider again a European call option where the stock price is USD 100, the strike price is USD 105, the risk-free rate is 4%, the volatility is 25%, and the time to maturity is one year. In this case:

$$d_1 = \frac{\ln(100/105) + (0.04 + 0.25^2/2) \times 1}{0.25\sqrt{1}} = 0.0898$$

From Equation (16.1):

$$\Delta = N(0.0898) = 0.5358$$

The trader who has sold 1 million options (i.e., has a short position in 1 million options) has a delta of:

$$(-1{,}000{,}000) \times 0.5358 = -535{,}800$$

To hedge the position, he or she should buy 535,800 shares. The gain (or loss) on the option will then be offset by the loss (or gain) on the shares. Suppose the stock price increases by USD 0.2, we would then expect the option price to increase by about USD 0.1072 (= 0.2 × 0.5358). The loss on the short option position will be

$$1{,}000{,}000 \times 0.1072 = 107{,}200$$

The gain on the 535,800 shares will be

$$535{,}800 \times 0.2 = 107{,}200$$

so that the net gain/loss is zero. One problem for the trader is that delta does not remain constant. Delta hedging is therefore not a "hedge and forget" strategy. The hedge must be adjusted every day (or even more frequently).[4] Adjustments to the hedge are referred to as *rebalancing.*

Suppose that after one day the stock price rises to USD 100.4. The option now has 251 days to maturity and the new delta is

$$N\left(\frac{\ln(100.4/105) + (0.04 + 0.25^2/2) \times 251/252}{0.25\sqrt{251/252}}\right) = 0.5418$$

The delta of the trader's position is now:

$$-1{,}000{,}000 \times 0.5418 = -541{,}800$$

so that the number of shares that must be bought for the hedge is now 541,800. The trader must therefore buy an additional 6,000 shares (= 541,800 − 535,800).

Suppose that the stock price falls to USD 99.9 by the end of the next day. The time to maturity is now 250 days and the new delta is

$$N\left(\frac{\ln(99.9/105) + (0.04 + 0.25^2/2) \times 250/252}{0.25\sqrt{250/252}}\right) = 0.5334$$

To rebalance the hedge, the trader must sell 8,400 shares to bring the holding down to 533,400 shares.

Delta hedging continues in this way until the option expires. If the stock price is above USD 105 as the end of the life of the option approaches, the delta approaches one and the trader is fully covered. (For example, if the stock price is USD 110 with two business days remaining, the delta is 0.9827.) If the stock price is below USD 105 as the end of the life of the option approaches, delta approaches zero.[5] For example, if the stock price is USD 100 with two business days remaining, delta is 0.0152.

Each day during the life of the option, the trader is well hedged and (when the Black-Scholes-Merton assumptions hold) the total cost of hedging the option day-by-day (after discounting) is usually close to its theoretical price.[6]

The hedging strategy can be compared to the stop-loss strategy we considered earlier. The trader does not choose to fully cover or go completely naked at any given time. Instead the position is partially covered. The extent to which the position is covered at any given time depends on delta, which can be thought of as a measure of the probability of the option being exercised.

The costs in delta hedging arise from the fact that the trader is always buying shares immediately after a price rise and selling immediately after a price fall. It is a "buy-high, sell-low" strategy, which is almost guaranteed to be costly. However, it has the effect of hedging the trader's position day-by-day.

16.3 VEGA

In all our calculations so far, we have assumed volatility remains constant. In practice, volatility changes through time, and this increases the risk for derivatives traders.

Consider the options trader in the previous section. Suppose that while the stock price stays at USD 100 during the first day of the option's life, the volatility increases from 25% to 28%. The trader's position would then decline in value since the options sold have increased in value. The Black-Scholes-Merton model gives the new option value as USD 10.75. This means that the options, which were sold for USD 10 million, are now worth USD 10.75 million. This is USD 1.19 million (= 10.75 − 9.56) more than they were worth before the volatility increased. Even though the stock price has not changed, the trader has lost money.

Vega is the Greek letter that measures the trader's exposure to volatility.[7] It is defined as:[8]

$$\text{vega} = \frac{\delta f}{\delta \sigma}$$

[4] In Chapter 14, we saw that delta is different in different parts of a tree, showing that delta hedging involves periodic adjustments to the hedge.

[5] The worst position for a trader is that the stock price is very close to the strike price as the maturity of the option approaches. Delta is then close to 0.5. If the option closes in-the-money, the trader is likely to be under-hedged at maturity; if the option closes out-of-the-money the trader is likely to be over-hedged at maturity.

[6] This can be verified with a Monte Carlo simulation.

[7] Vega, although it is referred to as a "Greek letter" is not one of the letters in the Greek alphabet.

[8] To be more precise, we can use calculus notation and write vega $= \frac{\partial f}{\partial \sigma}$.

where $\delta\sigma$ is a small change in volatility (with everything else remaining the same), and δf is the resultant change in the value of the derivative. The volatility σ is actually the implied volatility. (As mentioned earlier, σ is set equal to the implied volatility when Greek letters are calculated.)

A long position in either a call option or a put option has a positive vega. The vega of a European call (or put) option on a non-dividend paying stock is given by:

$$\text{vega} = S_0\sqrt{T}N'(d_1)$$

Here $N'(x)$ is the standard normal probability density function:[9]

$$N'(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

Vega is greatest for options that are close-to-the-money (i.e., options where the stock price is close to the strike price). As illustrated in Figure 16.4, vega tends to zero as the option moves deep-out-of-the-money or deep-in-the-money. When an option is deep-out-of-the-money, it almost certainly will not be exercised regardless of volatility changes; when it is deep-in-the-money, it will almost certainly be exercised regardless of volatility changes.

Consider again a call option on a stock worth USD 100 when the strike price is USD 105, the risk-free rate is 4%, the volatility is 25%, and the time to maturity is one year. As noted earlier, $d_1 = 0.0898$ and:

$$N'(d_1) = \frac{1}{\sqrt{2\pi}}e^{-0.0898^2/2} = 0.3973$$

This means that the vega of a long position in one option is

$$\text{vega} = 100 \times \sqrt{1} \times 0.3973 = 39.73$$

In the formula for vega, it is assumed that volatility is expressed as a decimal. The change in the price of the option for a 0.01 (1%) change in volatility is therefore USD 0.3973 (= 39.73 × 0.01). The change in the value of a portfolio that has sold options on 1 million shares is therefore USD −397,300 per 1% increase in volatility.

Note that vega suggests the 3% increase in volatility we considered earlier (i.e., from 25% to 28%) leads to a loss of USD 1,191,900 (= 3 × 397,300). This is approximately the same as the USD 1.19 million loss we calculated earlier by substituting a volatility of 28% for 25% into the Black-Scholes-Merton model. This illustrates that the value of an option is approximately linear in volatility.

[9] It is the derivative of $N(x)$ with respect to x.

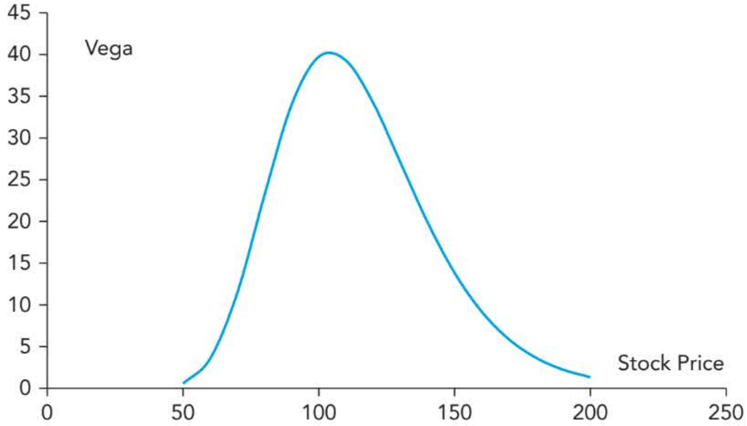


Figure 16.4 Variation of vega of an option with a stock price for a call or put option with a strike price of 105 and a time to maturity of one year when the risk-free rate is 4% and the volatility is 25%.

Calculating vega from the Black-Scholes-Merton model may seem strange since the model assumes that volatility is constant. Indeed, models have been developed where the volatility changes in an uncertain way over time.[10] However, the vega calculated from these models is similar to the Black-Scholes-Merton vega.

From a practical perspective, it makes sense for traders to monitor vega. If we assume interest rates are constant, the price of an option is a function of its implied volatility and the stock price.[11] If a trader can fully hedge against movements in stock price and implied volatility, there is very little risk.

Unfortunately, hedging vega risk is not as easy as hedging delta risk. The vega of a position in the underlying asset is zero. This means that trading the underlying asset does not affect the vega of a portfolio of derivatives dependent on the asset. Vega can only be adjusted by taking a position in another derivative dependent on the same asset.

Suppose a delta-neutral portfolio dependent on an asset price has a vega of USD −500. An option on the same asset can be traded that has a delta of USD 3.0 and a vega of USD 2.5. A trader can hedge vega by buying 200 of these options. The vega is reduced to zero because:

$$-500 + 200 \times 2.5 = 0$$

[10] See, for example, J. Hull and A. White, "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance*, 42 (June 1987): 281–300; and S. L. Heston, "A Closed Form Solution to Options on Assets with Stochastic Volatility with Application to Bonds and Currency Options," *Review of Financial Studies, 6, 2* (1993): 327–343.

[11] This is exactly true from the definition of implied volatility. The function is the Black-Scholes-Merton function.

However, an amount of delta equal to USD 600 (= 3.0 × 200) is introduced into the portfolio. The trader must therefore short 600 shares to maintain delta neutrality.

16.4 GAMMA

The gamma of a stock price–dependent derivative measures the sensitivity of its delta to the stock price. It is defined as:[12]

$$\text{gamma} = \frac{\delta(\Delta)}{\delta S}$$

where δS is a small change in the stock price, and $\delta(\Delta)$ is the resultant change in delta.

As Figure 16.1 illustrates, delta hedging assumes the relationship between the option price and the stock price is linear when it is actually non-linear. Gamma is a measure of the error made by this linearity assumption.

Consider Figure 16.5. When the stock price moves from S to S′, delta hedging assumes the option price will move from C to C′; in fact, it moves from C to C″.

Suppose the position in the stock is changed daily to maintain delta neutrality. The risk associated with stock price movements then depends on how much the stock price could change over one day, and how much curvature there is in the relationship between the option price and the stock price. Gamma measures this curvature. For a delta-neutral portfolio, gamma measures the risk associated with relatively large changes in stock price between hedge rebalancing. Small stock price changes do not create a lot of risk since the relationship between the option price and the stock price is approximately linear for small changes.[13]

As we show in the appendix, the change in the value of a portfolio when an option position is delta-neutral and the stock price changes by ΔS is approximately:

$$\frac{1}{2} \times \text{gamma} \times (\Delta S)^2$$

For example, suppose that the gamma of a short option position is −0.1, and the stock price increases by USD 2. The value of the option position can then be expected to decrease by 0.2.

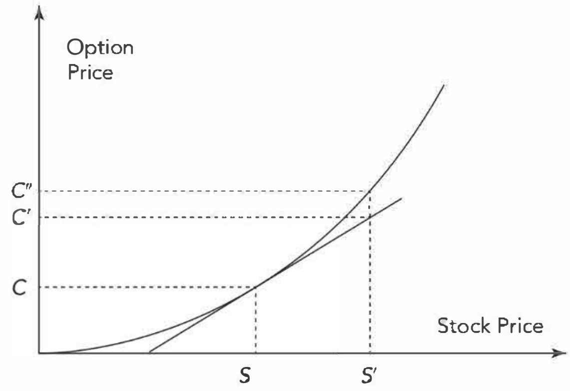


Figure 16.5 Hedging error introduced by non-linearity.

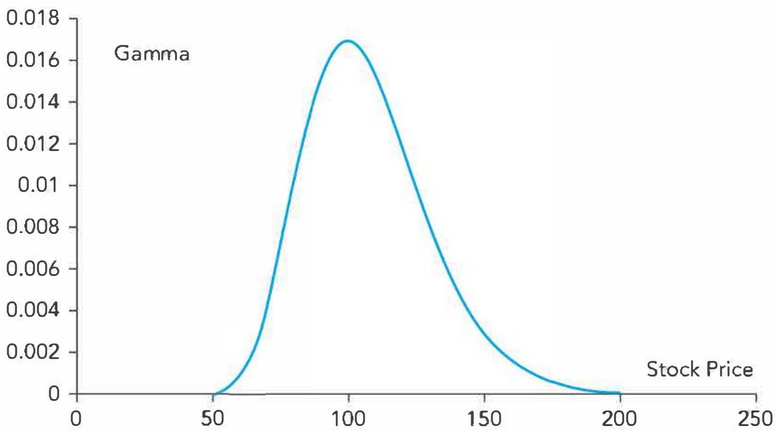


Figure 16.6 Variation of gamma of an option with stock price when the strike price is 105, the risk-free rate is 4%, the volatility is 25%, and the time to maturity is one year.

Gamma is positive for a long position in either a call option or put option. The gamma of a call (or put) option on a non-dividend paying stock is

$$\text{gamma} = \frac{N'(d_1)}{S\sigma\sqrt{T}}$$

Gamma is similar to vega in that it is greatest for options that are close-to-the-money and approaches zero as the option moves deep-in-the-money or deep-out-of-the-money. This is illustrated in Figure 16.6. One important difference between gamma and vega is that while vega increases as the time to maturity increases, gamma decreases.

Consider again a call option on a stock with a price of USD 100 when the strike price is USD 105, the time to maturity is one year, the risk-free rate is 4%, and the volatility is 25%. Because $d_1 = 0.0898$:

$$\text{gamma} = \frac{e^{-0.0898^2/2}/\sqrt{2\pi}}{100 \times 0.25 \times \sqrt{1}} = 0.0159$$

[12] To be more precise, we can use calculus notation and write $\text{gamma} = \frac{\partial^2 f}{\partial S^2}$, where f is the price of the derivative.

[13] Delta and gamma are analogous to the duration and convexity measurements that we introduced for interest rate changes in Chapter 12.

This means delta changes by 0.0159 per unit change in the stock price. We saw when we were considering delta hedging that, when the stock price increases from USD 100 to USD 100.4, delta increases by 0.0060 (= 0.5418 − 0.5358). Similarly, gamma suggests that the increase should be 0.0064 (= 0.4 × 0.0159).

Gamma (like vega) can be changed only by taking a position in a derivative. This is because the position in the underlying asset has zero gamma. The procedure for making gamma zero (by trading another option) is similar to that given earlier for making vega zero.

Suppose we want to make both gamma and vega zero. In theory, this can be done with two options. Suppose that the situation is as indicated in Table 16.1. The vega and gamma of the portfolio can be reduced to zero if positions x_A and x_B are taken in options A and B where:

$$600 + 2.5x_A + 4.5x_B = 0$$

and

$$36 + 0.05x_A + 0.15x_B = 0$$

The solutions to these equations are $x_A = 480$ and $x_B = -400$. We can therefore make vega and gamma zero by buying 480 of option A and selling 400 of option B. This introduces an amount of delta into the portfolio equal to:

$$480 \times 3 - 400 \times 4 = -160$$

It is therefore necessary for the trader to purchase 160 shares to maintain delta neutrality.

Table 16.1 Current Portfolio and Two Options that could be used to Hedge it

	Delta	Vega	Gamma
Current portfolio	0	600	36
Option A	3.0	2.5	0.05
Option B	4.0	4.5	0.15

16.5 THETA

The theta of an option is the rate of change in its value over time. When time is measured in years, the theta of long positions in European call options and put options are

$$\text{theta(call)} = -\frac{S_o N'(d_1)\sigma}{2\sqrt{T}} - rKe^{-rT}N(d_2)$$

$$\text{theta(put)} = -\frac{S_o N'(d_1)\sigma}{2\sqrt{T}} + rKe^{-rT}N(-d_2)$$

Consider again a European call option where the stock price is USD 100, the strike price is USD 105, the risk-free rate is 4%, the volatility is 25%, and the time to maturity is one year. The theta of the option is

$$\text{theta(call)} = -\frac{100 \times N'(0.08984) \times 0.25}{2\sqrt{1}} - 0.04 \times 105 \times e^{-0.04\times 1} \times N(-0.16016) = -6.73$$

This means that if nothing changes, a long position in the option loses money at the rate of USD 6.73 per year.

In practice, theta is usually quoted as a rate per day. Assuming there are 252 trading days in a year, a long position in the option loses money at the rate of USD 0.0267 (= 6.73/252) per day.

The Black-Scholes-Merton analysis can be used to show that theta is related to delta and gamma:

$$\text{theta} + rS_0 \times \text{delta} + \frac{1}{2}\sigma^2 S_0^2 \times \text{gamma} = rf \tag{16.2}$$

where f is the option value.

Theta is usually negative for a long option position because (with all else remaining constant) options lose value as time passes. Theta differs from the other Greek letters in that there is no uncertainty about the passage of time (whereas there is uncertainty about a stock's price or its volatility).

Despite this certainty regarding time, traders do like to monitor theta. One reason for this may be that Equation (16.2) shows that theta and gamma are negatively related for a delta-neutral portfolio. Theta therefore contains information about gamma when delta is maintained at zero. When theta is highly negative, gamma tends to be highly positive; when theta is highly positive, gamma tends to be highly negative.

16.6 RHO

The rho of an option measures its sensitivity to interest rates. When interest rates are expressed as decimals, the formulas for the rho of a European call and put option on a non-dividend paying stock are

$$\text{rho(call)} = KTe^{-rT}N(d_2)$$

$$\text{rho(put)} = -KTe^{-rT}N(-d_2)$$

Interest rate uncertainty is usually not as important for an option book as the uncertainty surrounding the asset price or its volatility.[14]

Consider again a European call option where the stock price is USD 100, the strike price is USD 105, the risk-free rate is 4%, the

[14] An exception here is, of course, interest rate options where interest rate uncertainty is of paramount importance.

volatility is 25%, and the time to maturity is one year. The rho of the option is

$$\text{rho(call)} = 105 \times 1 \times e^{-0.04 \times 1} \times N(-0.16016) = 44$$

This means that a 10-basis point (0.1%) increase in the interest rate (with nothing else changing) can be expected to increase the option price by USD 0.044 (= 0.001 × 44).

16.7 PORTFOLIOS

Any of the Greek letters for a portfolio of derivatives dependent on the same asset can be calculated as the weighted sum of the Greek letters for each portfolio component. For example, if a portfolio consists of the following items.

- A long position in 50,000 call options where the delta, vega, and gamma of each option are 0.46, 3.3, and 0.13 (respectively).
- A short position in 20,000 call options where the delta, vega, and gamma of each option are 0.33, 4.2, and 0.15 (respectively).
- A short position in 30,000 put options where the delta, vega, and gamma of each option are −0.54, 3.0, and 0.08 (respectively).

For the portfolio, the delta, vega, and gamma are

$$\begin{aligned} \text{delta} &= 50{,}000 \times 0.46 - 20{,}000 \times 0.33 - 30{,}000 \times (-0.54) \\ &= 32{,}600 \\ \text{vega} &= 50{,}000 \times 3.3 - 20{,}000 \times 4.2 - 30{,}000 \times 3.0 = -9{,}000 \\ \text{gamma} &= 50{,}000 \times 0.13 - 20{,}000 \times 0.15 - 30{,}000 \times 0.08 \\ &= 1{,}100 \end{aligned}$$

16.8 TRADING LIMITS

Traders working for banks and other large financial institutions are subject to limits on delta, vega, and gamma. There are several ways of expressing these limits.

For example, the delta limit might be expressed as USD 100,000. This amount is the derivatives exposure expressed as an equivalent position in the underlying asset. If the price of the underlying asset is USD 20, for example, the limit on delta, as we have calculated it in this chapter, would be USD 5,000 (= 100,000/20). Meanwhile, the vega limit might be set as USD 200,000 per 1%. This means that a 1% change in volatility can lead to a change of no more than USD 200,000 in the value of the portfolio. Finally, the gamma limit might be expressed as 500 deltas. This indicates that a USD 1 change in the price of the underlying asset can result in a change in delta of no more than 500.

Whereas delta is typically adjusted at the end of every day, it is too expensive to do the same for gamma and vega. This is because, while delta can be adjusted relatively easily by trading the underlying asset at a competitive price, it can be expensive to trade the derivatives necessary to adjust vega or gamma. If available, these derivatives are liable to be expensive.

As a result, traders must opportunistically look for ways that vega and gamma risk can be hedged. Luckily, there is a tendency for vega and gamma to disappear as time passes. Recall that Figures 16.4 and 16.6 show that vega and gamma are greatest for an option that is close-to-the-money and smallest for options that are deep-in or deep-out-of-the-money. While an option is often close-to-the-money when it is first sold, the price of the underlying asset can move so that the option is either deep-in or deep-out-of-the-money. As this happens, gamma and vega shrink. Of course, there is no guarantee that this will happen, and these two Greeks give rise to the biggest risks when options stay close-to-the-money.

Derivatives trading is also subject to economies of scale.[15] Consider two traders: one with a portfolio of ten options on an asset and another with a portfolio of 1,000 options on the same asset. Both traders can manage delta at the end of a day with a single trade by either buying or selling the underlying asset. However, while the second trader can offset the daily bid-ask spreads (one of the main transaction costs) using the profits on 1,000 derivatives transactions, the first trader's profits on just ten transactions may well be insufficient to cover these expenses. It is therefore not surprising that the derivatives business has traditionally been dominated by only a few large financial institutions.

Traders sometimes calculate several other Greek letters. These include *vanna* (the sensitivity of delta to volatility), *charm* (the sensitivity of delta to the passage of time), and *vomma* (the sensitivity of vega to a change in implied volatility). The Greek letters can be viewed in the context of a Taylor series expansion of the change in the value of a portfolio (as shown in the appendix).

The Greek letters look at relatively small changes in one variable over a short period of time. Traders may also carry out a sensitivity analysis examining the impact of changes to several variables over longer periods of time. For example, they could examine what would happen over the next month if there were a large increase or decrease in the price of the underlying asset combined with a change in volatility.

16.9 OTHER EUROPEAN OPTIONS

We have thus far provided formulas for delta, vega, gamma, theta, and rho for European options on non-dividend paying stocks. Table 16.2 shows how these formulas are

[15] Economies of scale refers to when the marginal cost of an activity decreases as the amount of the activity increases.

Table 16.2 Greek Letters for European Options on An Asset Providing a Dividend Yield at Rate q

Greek Letter	Call Option	Put Option
Delta	$e^{-qT}N(d_1)$	$e^{-qT}[N(d_1) - 1]$
Vega	$S_0\sqrt{T}N'(d_1)e^{-qT}$	$S_0\sqrt{T}N'(d_1)e^{-qT}$
Gamma	$\dfrac{N'(d_1)e^{-qT}}{S_0\sigma\sqrt{T}}$	$\dfrac{N'(d_1)e^{-qT}}{S_0\sigma\sqrt{T}}$
Theta	$-\dfrac{S_0N'(d_1)\sigma e^{-qT}}{(2\sqrt{T})} + qS_0N(d_1)e^{-qT} - rKN(d_2)e^{-rT}$	$-\dfrac{S_0N'(d_1)\sigma e^{-qT}}{(2\sqrt{T})} - qS_0N(-d_1)e^{-qT} + rKN(-d_2)e^{-rT}$
Rho	$KTe^{-rT}N(d_2)$	$-KTe^{-rT}N(-d_2)$

adjusted when there is a dividend yield at rate q. For options on stock indices, we set q equal to the dividend yield on the index. For options on currencies, we set q equal to the foreign risk-free rate. For options on futures, we set $q = r$. An exception is the formula for the rho of a European futures option; this is $-fT$, where f is the price of the option.

As in Chapter 15, the formulas for d_1 and d_2 in Table 16.2 are adjusted to:

$$d_1 = \frac{ln(S_0/K) + (r - q + \sigma^2/2)T}{\sigma\sqrt{T}}$$

$$d_2 = \frac{ln(S_0/K) + (r - q - \sigma^2/2)T}{\sigma\sqrt{T}}$$

16.10 AMERICAN OPTIONS

Binomial trees can be used to calculate Greek letters for American options. Figure 16.7 shows the binomial tree constructed in Figure 14.8 of Chapter 14 to value a call option on a currency.

We calculate delta (as in Chapter 14) using the two nodes at the end of the first time step.[16] It is the change in the option price divided by the change in the exchange rate:

$$\text{delta} = \frac{0.0394 - 0.0051}{0.8282 - 0.7346} = 0.37$$

To obtain gamma we calculate the two deltas at the end of the second step. The delta calculated from the upper two nodes is

$$\text{delta} = \frac{0.0794 - 0.0127}{0.8794 - 0.7800} = 0.6710$$

This delta can be assumed to apply to the average of the two upper exchange rates: 0.8297 (= (0.8794 + 0.7800)/2). The delta calculated from the lower two nodes is

$$\text{delta} = \frac{0.0127 - 0}{0.7800 - 0.6918} = 0.1440$$

This delta can be assumed to apply to the average of the two lower exchange rates: 0.7359 (= (0.7800 + 0.6918)/2). Gamma is the change in delta divided by the change in the exchange rate:[17]

$$\text{gamma} = \frac{0.6710 - 0.1440}{0.8297 - 0.7359} = 5.6$$

Theta can be calculated by comparing the prices at the initial node and the middle node at the end of the second time step. We see that the option price has declined by 0.0061 (= 0.0188 − 0.0127) over the six-month period between the nodes. Theta is therefore −0.012 (−0.0061/0.5) per year.

To calculate vega, we must make a small change to volatility and construct a new tree. The number of time steps is kept the same. When we construct a new tree for the example in Figure 16.7 (by setting the volatility equal to 12.1% instead of 12%), we find that the value of the option increases by 0.00026 to become 0.01904. Vega is therefore 0.0026 per 1% change in volatility. Rho can be calculated similarly by increasing the interest rate and constructing a new tree.

[16] In theory, this gives delta at the end of the first time step not at time zero. For a large tree, this makes very little difference. For more accuracy, the tree could be started at time $-\Delta t$ so that there are two nodes at time zero.

[17] In theory, this gives delta at the end of the second time step, not at time zero. For a large tree, this makes very little difference. For more accuracy, the tree could be started at time $-2\Delta t$ so that there are two nodes at time zero.

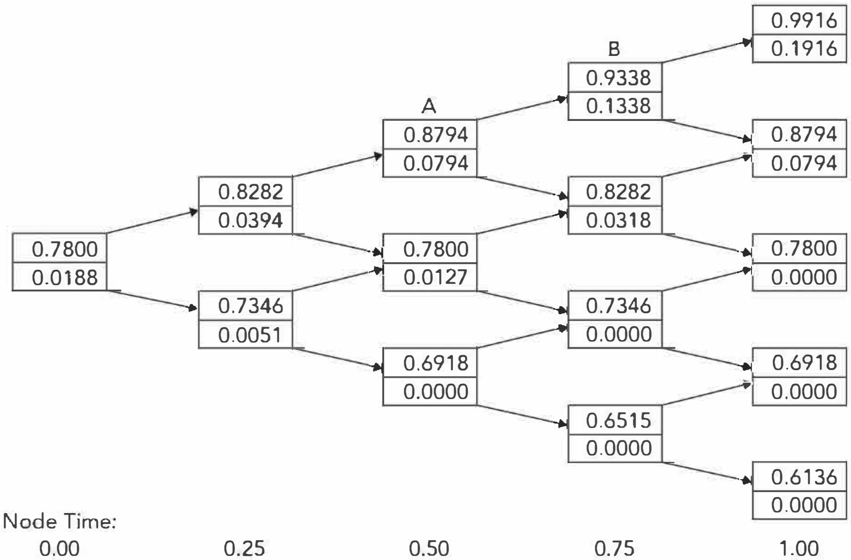


Figure 16.7 **Shown is a four-step tree for valuing an American call option on a currency with a strike price of 0.8000. The current exchange rate is 0.7800, the volatility is 12%, the domestic risk-free rate is 2%, the foreign risk-free rate is 6%, and the time to maturity is one year. The upper number at each node is the stock price, and the lower number is the option price.**

16.11 PORTFOLIO INSURANCE

The delta hedging procedure described earlier in this chapter involves creating the opposite position to that being hedged. In our example, a trader wished to hedge a short position in a one-year option with a strike price of USD 105 and a life of one year. The effect of the day-to-day delta hedging procedure we described is to approximately create a long position in the call option.

Asset managers are sometimes interested in acquiring a put option on their portfolio to provide protection against a market decline. They can create the option synthetically by imagining they are hedging a short position in the put option. Instead of taking a position in the underlying asset that neutralizes the delta of an existing option, they match the delta of the option they are trying to create.

Suppose a portfolio is worth USD 100 million and that the asset manager wants to provide put-option protection against the value falling below USD 90 million during the next six months. The portfolio has a volatility of 25%, the risk-free rate is 3%, and the dividend yield is 2%. The manager therefore wants to create a put option where $S_0 = 100$ million, $K = 90$ million, $r = 0.03$, $q = 0.02$, $\sigma = 0.25$, and $T = 0.5$. From Table 16.2, the delta of the option is $e^{-qT}[N(d_1)-1]$. In this case:

$$d_1 = \frac{ln(100/90) + (0.03 - 0.02 + 0.25^2/2) \times 0.5}{0.25\sqrt{0.5}} = 0.7127$$

and the delta is

$$e^{-0.02\times 0.5}[N(0.7127) - 1] = -0.236$$

To match deltas, the asset manager should therefore sell 23.6% of the portfolio. As time passes, delta changes. If the asset manager ensures that he or she has always sold an amount of the original portfolio to match the delta, the required option will be approximately created. There will be a cost associated with this strategy, however, since the manager will need to sell the portfolio in a declining market and buy it back in a rising market.

If the portfolio mirrors an index, the portfolio manager can keep the portfolio intact and buy options on index futures on an exchange. If the options market is not liquid enough to accommodate the manager's desired trade, the required options on index futures can be created synthetically by trading the futures contracts themselves in the way we have described.

Portfolio insurance was popular up until the late 1980s. However, its weaknesses were exposed when the Dow Jones Industrial Average dropped by more than 20% over a single day (this was Monday, October 19, 1987, and it is commonly referred to as Black Monday).

When the market started to drop, portfolio insurers sent sell orders to the exchange. This, in turn, made the market drop even further. As a result, there were more sell orders from portfolio insurers. The volume of sell orders overloaded the stock exchanges, and trades could not be handled immediately. As a result, portfolio insurance proved less effective than predicted. Unsurprisingly, it has become much less popular since 1987.

SUMMARY

Greek letters measure different aspects of risk in derivatives portfolios. Delta measures the sensitivity of a portfolio's value to changes in the price of the underlying asset. Vega measures the sensitivity of a portfolio's value to the implied volatility of the underlying asset. Gamma measures the sensitivity of a portfolio's delta to changes in the price of the underlying asset. Theta measures the sensitivity of the portfolio's value to the passage of time. Rho measures the sensitivity of the portfolio's value to changes in the level of interest rates. In all cases, it is assumed that one variable is changed with all others remaining constant.

In practice, traders rebalance their portfolios (by trading the underlying asset) at least once a day to bring delta close to zero. Changing vega or gamma requires the use of derivatives trades. These can be expensive and, therefore, vega and gamma are rebalanced much less frequently than delta.

Theta can be calculated from delta and gamma. There are analytic formulas for calculating the Greek letters for European options. Binomial trees can be used to calculate Greek letters for American options numerically.

Delta can be used to create an option. A position is maintained in the underlying asset to match the delta of the option required. This strategy was once a popular approach for portfolio managers who wanted to create synthetic put options on their portfolios. It has become less popular because it did not work as well as expected during the market crash of 1987.

APPENDIX

Taylor Series Expansion for Change in Value of a Portfolio

Consider a portfolio of derivatives dependent on a particular asset. The value of the portfolio is a function of the asset price, the volatility of the asset price, and time t. (This assumes interest rates and dividend yield do not change.) A Taylor series expansion gives the change in the portfolio value, P, in a short period of time Δt as:

$$\Delta P = \frac{\partial P}{\partial S}\Delta S + \frac{\partial P}{\partial \sigma}\Delta \sigma + \frac{\partial P}{\partial t}\Delta t + \frac{1}{2}\frac{\partial^2 P}{\partial S^2}(\Delta S)^2 + \cdots$$

This can be written

$$\Delta P = \text{delta} \times \Delta S + \text{vega} \times \Delta \sigma + \text{theta} \times \Delta t + \frac{1}{2} \times \text{gamma} \times (\Delta S)^2 + \cdots$$

The hedging procedures we have discussed in this chapter involve making the Greek letters in this equation zero. Delta hedging deals with the first term on the right-hand side; vega hedging deals with the second term. Other Greek letters, which have been briefly mentioned, such as vanna, charm, and vomma can be used to hedge higher order terms in the expansion. (These higher order terms are not shown in the previous expansion).

If a portfolio has a delta of zero, and the asset price changes without any time passing or any change to the volatility, the Taylor series expansion shows that it is approximately true that

$$\Delta P = \frac{1}{2} \times \text{gamma} \times (\Delta S)^2$$

This is the result in section 16.4.

The following questions are intended to help candidates understand the material. They are not actual FRM exam questions.

QUESTIONS

Short Concept Questions

16.1 Explain the stop-loss strategy for hedging a short position in a call option. What is its drawback?

16.2 The delta for a portfolio of derivatives dependent on a particular asset is −400. What does this mean?

16.3 "Delta hedging a short position in a call option is a 'buy-high, sell-low' trading strategy." Explain this statement.

16.4 What is the sign of the delta of a long position in (a) a call option and (b) a put option?

16.5 What is the advantage to a trader of bringing the vega of a portfolio close to zero?

16.6 How does vega vary with stock price for a European put option?

16.7 What is the advantage to a trader of making (a) delta and (b) gamma zero?

16.8 How are (a) theta and (b) rho defined?

16.9 Why are there economies of scale in trading derivatives?

16.10 Explain why portfolio insurance may have played a part in the crash of 1987.

Practice Questions

16.11 An out-of-the-money option with a strike price of 30 has a theoretical price of USD 4. A trader hedges the option by buying the stock at USD 30.1 and selling at USD 29.9. How many times would the stock need to be traded to equal the cost of the option? (Ignore the impact of discounting on the present value of future costs.)

16.12 What is the delta of a short position on 100,000 call options on a stock with a market price and strike price of USD 40 when the risk-free rate is 5%, the volatility is 22%, and the time to maturity is nine months?

16.13 In Question 16.12, what trade should be done to create a delta-neutral position? (Assume that the trader has no other positions dependent on the stock price.) If the stock price increases to USD 41 within a very short period, what further trade is necessary?

16.14 Suppose that the vega of a portfolio of options dependent on a particular asset is USD 4.5 per 1%. The volatility of the asset decreases from 23% to 21%. What do you expect to happen to the value of the portfolio?

16.15 What is the vega of a European put option on a stock index when the index level is USD 1,500, the strike price is USD 1,400, the risk-free rate is 5%, the dividend yield is 2%, the volatility is 18%, and the time to maturity is three months. How can this be interpreted?

16.16 A delta-neutral portfolio has a gamma of −20. The price of the underlying asset suddenly increases by USD 3. Estimate what happens to the value of the portfolio. What difference does it make if the price of the underlying asset suddenly decreases by USD 3?

16.17 From the information in the following table, estimate (a) what position should be taken in option A and the underlying asset for vega and delta neutrality, and (b) what position should be taken in option B and the underlying asset for gamma and delta neutrality. Note: when answering part (b) do not assume that the position in part (a) has been taken.

	Delta	Vega	Gamma
Portfolio	0	400	60
Option A	0.8	2	0.4
Option B	−0.6	3	0.5

16.18 In Question 16.17, what position should be taken in the two options and the underlying asset for delta, vega, and gamma neutrality?

16.19 Use the binomial tree in Figure 14.7 of Chapter 14 to estimate the delta, gamma, and theta of the option.

16.20 A trader wants to create synthetically a nine-month European put futures option on 1 million times an index. The futures price is USD 2,500, the strike price is USD 2,400, the risk-free rate is 2%, and the volatility of the futures price is 20%. What position should the trader take in futures contracts initially? How does this differ from the position the trader would take if he or she were hedging the same nine-month European put futures option on 1 million times the index?

ANSWERS

Short Concept Questions

16.1 The stop-loss strategy involves covering the option by buying the asset that will have to be delivered as soon as the option moves in-the-money and selling when it moves out-of-the-money. It works well for some scenarios and badly for others. It is not a hedging scheme where the present value of the cost of hedging the option is always approximately equal to its theoretical price.

16.2 When the asset price increases by USD 1, the value of the portfolio declines by USD 400. When the asset price decreases by USD 1, the value of the portfolio increases by USD 400.

16.3 Delta hedging involves buying Δ shares for each option sold, where Δ is the delta of a long position in the option. As the share price rises, delta increases and more shares have to be bought. As the share price falls, delta decreases and shares have to be sold. The trader is therefore always buying after a price increase and selling after a price decrease.

16.4 The delta of a long position in a call option is positive. The delta of a long position in a put option is negative.

16.5 This reduces to almost zero the trader's exposure to a change in volatility.

16.6 Vega is greatest for at-the-money options. It declines to zero as an option moves deep-in or deep-out-of-the-money.

16.7 A zero delta provides protection against small changes in the price of the underlying asset. A zero gamma provides protection against larger changes in the price of the underlying asset.

16.8 Theta measures the rate of change in the value of a portfolio when time passes with all else remaining the same. Rho measures the sensitivity of the portfolio to a change in the interest rate.

16.9 To change the delta of a portfolio of derivatives dependent on a particular asset, a single trade in the underlying asset is required. This is true regardless of the size of the portfolio.

16.10 Portfolio insurance involves creating a synthetic put option on the portfolio. This is accomplished by selling the portfolio in a declining market. If enough portfolio managers are doing this, it will have the effect of accentuating the decline, leading to more selling.

Solved Problems

16.11 There is a cost of USD 0.1 each time the stock is bought or sold. The total expected cost of hedging the option should be the theoretical price of USD 4.0. We therefore expect buying or selling to take place roughly 4.0/0.2 = 20 times.

16.12 The delta of a long position in one option is $N(d_1)$. In this case:

$$d_1 = \frac{\ln(40/40) + (0.05 + 0.22^2/2) \times 0.75}{0.22\sqrt{0.75}} = 0.2921$$

so that $N(d_1) = 0.615$. The delta of a short position in one option is -0.615 and the delta of a short position in 100,000 options is $-61{,}500$.

16.13 The trader should buy 61,500 shares of the stock to create a delta-neutral position. If the stock price then moves up to USD 41:

$$d_1 = \frac{\ln(41/40) + (0.05 + 0.22^2/2) \times 0.75}{0.22\sqrt{0.75}} = 0.4217$$

and $N(d_1) = 0.663$. The delta of the option position is $-66{,}300$ and a further 4,800 shares should be purchased.

16.14 We expect the value of the portfolio to decrease by $2 \times 4.5 = 9.0$. (Although 2% is a relatively large change, options are close to linearly dependent on volatility, so we can expect the answer to be accurate.)

16.15 The vega is

$$S_0\sqrt{T}N'(d_1)e^{-qT}$$

In this case, $S_0 = 1{,}500$, $K = 1{,}400$, $r = 0.05$, $q = 0.02$, $\sigma = 0.18$, and $T = 0.25$.

$$d_1 = \frac{\ln(1{,}500/1{,}400) + (0.05 - 0.02 + 0.18^2/2) \times 0.25}{0.18\sqrt{0.25}} = 0.8949$$

and vega is

$$1{,}500 \times \sqrt{0.25} \times \frac{1}{\sqrt{2\pi}}e^{-0.8949^2/2} \times e^{-0.02 \times 0.25} = 199$$

This means that the value of a long position increases by $199 \times 0.01 = 1.99$ if volatility increases by 1% (= 0.01) from 18% to 19%. Similarly, it decreases by 1.99 if the volatility decreases from 18% to 17%.

16.16 The USD value of the portfolio will change by:

$$\frac{1}{2} \times (-20) \times 3^2 = -90$$

The change in the value of the portfolio is the same if the value of the underlying asset decreases by USD 3.

16.17 For vega neutrality, we can take a position of −200 in option A. This will create a delta of $0.8 \times (-200) = -160$, and 160 of the underlying asset should be purchased. For gamma neutrality, we can take a position of −120 in option B. This will create a delta of $-0.6 \times (-120) = 72$, and 72 of the underlying asset should be sold.

16.18 If the position is x_A in option A and x_B in option B we require

$$400 + 2x_A + 3x_B = 0$$
$$60 + 0.4x_A + 0.5x_B = 0$$

The solution to these equations is $x_A = 100$, $x_B = -200$. The position taken should therefore be a long position of 100 in option A and a short position of 200 in option B. This creates a delta of:

$$0.8 \times 100 + (-0.6) \times (-200) = 200$$

It is therefore necessary to sell 200 of the asset to maintain a delta of zero.

16.19 The delta estimate is

$$\frac{202.098 - 38.812}{2{,}657.878 - 2{,}351.500} = 0.5330$$

For gamma, we calculate two deltas from the nodes at the end of the second time step. The delta from the upper nodes, where the average index is 2,662.863, is

$$\frac{328.791 - 78.279}{2{,}825.726 - 2{,}500} = 0.7691$$

The delta calculated from the lower two nodes, where the average index is 2,355.911, is

$$\frac{78.279 - 0}{2{,}500 - 2{,}211.821} = 0.2716$$

The estimate of gamma is

$$\frac{0.7691 - 0.2716}{2{,}662.863 - 2{,}355.911} = 0.00162$$

The estimate of theta (per year) is

$$\frac{78.279 - 119.579}{0.3333} = -123.9$$

16.20 The delta of a long position in a put option on a futures price is $e^{-rT}[N(d_1) - 1]$. In this case:

$$d_1 = \frac{\ln(2{,}500/2{,}400) + (0.2^2/2) \times 0.75}{0.2\sqrt{0.75}} = 0.3223$$

and delta is

$$e^{-0.02 \times 0.75}[N(0.3223) - 1] = -0.368$$

The trader should short futures contracts on 368,000 times the index initially to match the delta of the position that is desired. If the trader were hedging 1 million put futures contracts he or she would take a long position in futures contracts on 368,000 times the index.

INDEX

A

B

C

D

E

F

N

O

P

R

S