CFA Institute®
CFA Program

FIXED INCOME, DERIVATIVES, ALTERNATIVE INVESTMENTS, AND PORTFOLIO MANAGEMENT

CFA® Program Curriculum
2022 • LEVEL I • VOLUME 5

ISBN 978-1-950157-46-4 (paper)
ISBN 978-1-950157-70-9 (ebk)

10 9 8 7 6 5 4 3 2 1

CONTENTS

Portfolio Management

How to Use the CFA Program Curriculum

Congratulations on your decision to enter the Chartered Financial Analyst (CFA®) Program. This exciting and rewarding program of study reflects your desire to become a serious investment professional. You are embarking on a program noted for its high ethical standards and the breadth of knowledge, skills, and abilities (competencies) it develops. Your commitment should be educationally and professionally rewarding.

The credential you seek is respected around the world as a mark of accomplishment and dedication. Each level of the program represents a distinct achievement in professional development. Successful completion of the program is rewarded with membership in a prestigious global community of investment professionals. CFA charterholders are dedicated to life-long learning and maintaining currency with the ever-changing dynamics of a challenging profession. CFA Program enrollment represents the first step toward a career-long commitment to professional education.

The CFA exam measures your mastery of the core knowledge, skills, and abilities required to succeed as an investment professional. These core competencies are the basis for the Candidate Body of Knowledge (CBOK™). The CBOK consists of four components:

- A broad outline that lists the major CFA Program topic areas (www.cfainstitute.org/programs/cfa/curriculum/cbok);
- Topic area weights that indicate the relative exam weightings of the top-level topic areas (www.cfainstitute.org/programs/cfa/curriculum);
- Learning outcome statements (LOS) that advise candidates about the specific knowledge, skills, and abilities they should acquire from readings covering a topic area (LOS are provided in candidate study sessions and at the beginning of each reading); and
- CFA Program curriculum that candidates receive upon exam registration.

Therefore, the key to your success on the CFA exams is studying and understanding the CBOK. The following sections provide background on the CBOK, the organization of the curriculum, features of the curriculum, and tips for designing an effective personal study program.

BACKGROUND ON THE CBOK

CFA Program is grounded in the practice of the investment profession. CFA Institute performs a continuous practice analysis with investment professionals around the world to determine the competencies that are relevant to the profession, beginning with the Global Body of Investment Knowledge (GBIK®). Regional expert panels and targeted surveys are conducted annually to verify and reinforce the continuous feedback about the GBIK. The practice analysis process ultimately defines the CBOK. The CBOK reflects the competencies that are generally accepted and applied by investment professionals. These competencies are used in practice in a generalist context and are expected to be demonstrated by a recently qualified CFA charterholder.

The CFA Institute staff—in conjunction with the Education Advisory Committee and Curriculum Level Advisors, who consist of practicing CFA charterholders—designs the CFA Program curriculum in order to deliver the CBOK to candidates. The exams, also written by CFA charterholders, are designed to allow you to demonstrate your mastery of the CBOK as set forth in the CFA Program curriculum. As you structure your personal study program, you should emphasize mastery of the CBOK and the practical application of that knowledge. For more information on the practice analysis, CBOK, and development of the CFA Program curriculum, please visit www.cfainstitute.org.

ORGANIZATION OF THE CURRICULUM

The Level I CFA Program curriculum is organized into 10 topic areas. Each topic area begins with a brief statement of the material and the depth of knowledge expected. It is then divided into one or more study sessions. These study sessions should form the basic structure of your reading and preparation. Each study session includes a statement of its structure and objective and is further divided into assigned readings. An outline illustrating the organization of these study sessions can be found at the front of each volume of the curriculum.

The readings are commissioned by CFA Institute and written by content experts, including investment professionals and university professors. Each reading includes LOS and the core material to be studied, often a combination of text, exhibits, and in-text examples and questions. End of Reading Questions (EORQs) followed by solutions help you understand and master the material. The LOS indicate what you should be able to accomplish after studying the material. The LOS, the core material, and the EORQs are dependent on each other, with the core material and EORQs providing context for understanding the scope of the LOS and enabling you to apply a principle or concept in a variety of scenarios.

The entire readings, including the EORQs, are the basis for all exam questions and are selected or developed specifically to teach the knowledge, skills, and abilities reflected in the CBOK.

You should use the LOS to guide and focus your study because each exam question is based on one or more LOS and the core material and practice problems associated with the LOS. As a candidate, you are responsible for the entirety of the required material in a study session.

We encourage you to review the information about the LOS on our website (www.cfainstitute.org/programs/cfa/curriculum/study-sessions), including the descriptions of LOS “command words” on the candidate resources page at www.cfainstitute.org.

FEATURES OF THE CURRICULUM

End of Reading Questions/Solutions *All End of Reading Questions (EORQs) as well as their solutions are part of the curriculum and are required material for the exam.* In addition to the in-text examples and questions, these EORQs help demonstrate practical applications and reinforce your understanding of the concepts presented. Some of these EORQs are adapted from past CFA exams and/or may serve as a basis for exam questions.

Glossary For your convenience, each volume includes a comprehensive Glossary. Throughout the curriculum, a **bolded** word in a reading denotes a term defined in the Glossary.

Note that the digital curriculum that is included in your exam registration fee is searchable for key words, including Glossary terms.

LOS Self-Check We have inserted checkboxes next to each LOS that you can use to track your progress in mastering the concepts in each reading.

Source Material The CFA Institute curriculum cites textbooks, journal articles, and other publications that provide additional context or information about topics covered in the readings. As a candidate, you are not responsible for familiarity with the original source materials cited in the curriculum.

Note that some readings may contain a web address or URL. The referenced sites were live at the time the reading was written or updated but may have been deactivated since then.

Some readings in the curriculum cite articles published in the *Financial Analysts Journal®*, which is the flagship publication of CFA Institute. Since its launch in 1945, the *Financial Analysts Journal* has established itself as the leading practitioner-oriented journal in the investment management community. Over the years, it has advanced the knowledge and understanding of the practice of investment management through the publication of peer-reviewed practitioner-relevant research from leading academics and practitioners. It has also featured thought-provoking opinion pieces that advance the common level of discourse within the investment management profession. Some of the most influential research in the area of investment management has appeared in the pages of the *Financial Analysts Journal*, and several Nobel laureates have contributed articles.

Candidates are not responsible for familiarity with *Financial Analysts Journal* articles that are cited in the curriculum. But, as your time and studies allow, we strongly encourage you to begin supplementing your understanding of key investment management issues by reading this, and other, CFA Institute practice-oriented publications through the Research & Analysis webpage (www.cfainstitute.org/en/research).

Errata The curriculum development process is rigorous and includes multiple rounds of reviews by content experts. Despite our efforts to produce a curriculum that is free of errors, there are times when we must make corrections. Curriculum errata are periodically updated and posted by exam level and test date online (www.cfainstitute.org/en/programs/submit-errata). If you believe you have found an error in the curriculum, you can submit your concerns through our curriculum errata reporting process found at the bottom of the Curriculum Errata webpage.

DESIGNING YOUR PERSONAL STUDY PROGRAM

Create a Schedule An orderly, systematic approach to exam preparation is critical. You should dedicate a consistent block of time every week to reading and studying. Complete all assigned readings and the associated problems and solutions in each study session. Review the LOS both before and after you study each reading to ensure that you have mastered the applicable content and can demonstrate the knowledge, skills, and abilities described by the LOS and the assigned reading. Use the LOS self-check to track your progress and highlight areas of weakness for later review.

Successful candidates report an average of more than 300 hours preparing for each exam. Your preparation time will vary based on your prior education and experience, and you will probably spend more time on some study sessions than on others.

You should allow ample time for both in-depth study of all topic areas and additional concentration on those topic areas for which you feel the least prepared.

CFA INSTITUTE LEARNING ECOSYSTEM (LES)

As you prepare for your exam, we will email you important exam updates, testing policies, and study tips. Be sure to read these carefully.

Your exam registration fee includes access to the CFA Program Learning Ecosystem (LES). This digital learning platform provides access, even offline, to all of the readings and End of Reading Questions found in the print curriculum organized as a series of shorter online lessons with associated EORQs. This tool is your one-stop location for all study materials, including practice questions and mock exams.

The LES provides the following supplemental study tools:

Structured and Adaptive Study Plans The LES offers two ways to plan your study through the curriculum. The first is a structured plan that allows you to move through the material in the way that you feel best suits your learning. The second is an adaptive study plan based on the results of an assessment test that uses actual practice questions.

Regardless of your chosen study path, the LES tracks your level of proficiency in each topic area and presents you with a dashboard of where you stand in terms of proficiency so that you can allocate your study time efficiently.

Flashcards and Game Center The LES offers all the Glossary terms as Flashcards and tracks correct and incorrect answers. Flashcards can be filtered both by curriculum topic area and by action taken—for example, answered correctly, unanswered, and so on. These Flashcards provide a flexible way to study Glossary item definitions.

The Game Center provides several engaging ways to interact with the Flashcards in a game context. Each game tests your knowledge of the Glossary terms a in different way. Your results are scored and presented, along with a summary of candidates with high scores on the game, on your Dashboard.

Discussion Board The Discussion Board within the LES provides a way for you to interact with other candidates as you pursue your study plan. Discussions can happen at the level of individual lessons to raise questions about material in those lessons that you or other candidates can clarify or comment on. Discussions can also be posted at the level of topics or in the initial Welcome section to connect with other candidates in your area.

Practice Question Bank The LES offers access to a question bank of hundreds of practice questions that are in addition to the End of Reading Questions. These practice questions, only available on the LES, are intended to help you assess your mastery of individual topic areas as you progress through your studies. After each practice question, you will receive immediate feedback noting the correct response and indicating the relevant assigned reading so you can identify areas of weakness for further study.

Mock Exams The LES also includes access to three-hour Mock Exams that simulate the morning and afternoon sessions of the actual CFA exam. These Mock Exams are intended to be taken after you complete your study of the full curriculum and take practice questions so you can test your understanding of the curriculum and your readiness for the exam. If you take these Mock Exams within the LES, you will receive feedback afterward that notes the correct responses and indicates the relevant assigned readings so you can assess areas of weakness for further study. We recommend that you take Mock Exams during the final stages of your preparation for the actual CFA exam. For more information on the Mock Exams, please visit www.cfainstitute.org.

PREP PROVIDERS

You may choose to seek study support outside CFA Institute in the form of exam prep providers. After your CFA Program enrollment, you may receive numerous solicitations for exam prep courses and review materials. When considering a prep course, make sure the provider is committed to following the CFA Institute guidelines and high standards in its offerings.

Remember, however, that there are no shortcuts to success on the CFA exams; reading and studying the CFA Program curriculum *is* the key to success on the exam. The CFA Program exams reference only the CFA Institute assigned curriculum; no prep course or review course materials are consulted or referenced.

SUMMARY

Every question on the CFA exam is based on the content contained in the required readings and on one or more LOS. Frequently, an exam question is based on a specific example highlighted within a reading or on a specific practice problem and its solution. To make effective use of the CFA Program curriculum, please remember these key points:

1. All pages of the curriculum are required reading for the exam.
2. All questions, problems, and their solutions are part of the curriculum and are required study material for the exam. These questions are found at the end of the readings in the print versions of the curriculum. In the LES, these questions appear directly after the lesson with which they are associated. The LES provides immediate feedback on your answers and tracks your performance on these questions throughout your study.
3. We strongly encourage you to use the CFA Program Learning Ecosystem. In addition to providing access to all the curriculum material, including EORQs, in the form of shorter, focused lessons, the LES offers structured and adaptive study planning, a Discussion Board to communicate with other candidates, Flashcards, a Game Center for study activities, a test bank of practice questions, and online Mock Exams. Other supplemental study tools, such as eBook and PDF versions of the print curriculum, and additional candidate resources are available at www.cfainstitute.org.
4. Using the study planner, create a schedule and commit sufficient study time to cover the study sessions. You should also plan to review the materials, answer practice questions, and take Mock Exams.
5. Some of the concepts in the study sessions may be superseded by updated rulings and/or pronouncements issued after a reading was published. Candidates are expected to be familiar with the overall analytical framework contained in the assigned readings. Candidates are not responsible for changes that occur after the material was written.

FEEDBACK

At CFA Institute, we are committed to delivering a comprehensive and rigorous curriculum for the development of competent, ethically grounded investment professionals. We rely on candidate and investment professional comments and feedback as we work to improve the curriculum, supplemental study tools, and candidate resources.

Please send any comments or feedback to info@cfainstitute.org. You can be assured that we will review your suggestions carefully. Ongoing improvements in the curriculum will help you prepare for success on the upcoming exams and for a lifetime of learning as a serious investment professional.

Fixed Income

STUDY SESSIONS

Study Session 13	Fixed Income (1)
Study Session 14	Fixed Income (2)

TOPIC LEVEL LEARNING OUTCOME

The candidate should be able to describe fixed-income securities and their markets, yield measures, risk factors, and valuation measures and drivers. The candidate should also be able to calculate yields and values of fixed-income securities.

Fixed-income securities continue to represent the largest capital market segment in the financial ecosystem and the primary means in which institutions, governments, and other issuers raise capital globally. Institutions and individuals use fixed-income investments in a wide range of applications including asset liability management, income generation, and principal preservation. Since the global financial crisis of 2008, evaluating risk—in particular, credit risk—for fixed-income securities has become an increasingly important aspect for this asset class.

FIXED INCOME

STUDY SESSION 14

Fixed Income (2)

This study session examines the fundamental elements underlying bond returns and risks with a specific focus on interest rate and credit risk. Duration, convexity, and other key measures for assessing a bond's sensitivity to interest rate risk are introduced. An explanation of credit risk and the use of credit analysis for risky bonds concludes the session.

READING ASSIGNMENTS

Reading 43	Understanding Fixed-Income Risk and Return by James F. Adams, PhD, CFA, and Donald J. Smith, PhD
Reading 44	Fundamentals of Credit Analysis by Christopher L. Gootkind, CFA

READING

43

Understanding Fixed-Income Risk and Return

by James F. Adams, PhD, CFA, and Donald J. Smith, PhD

James F. Adams, PhD, CFA, is at New York University (USA). Donald J. Smith, PhD, is at Boston University Questrom School of Business (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** calculate and interpret the sources of return from investing in a fixed-rate bond;
☐	**b.** define, calculate, and interpret Macaulay, modified, and effective durations;
☐	**c.** explain why effective duration is the most appropriate measure of interest rate risk for bonds with embedded options;
☐	**d.** define key rate duration and describe the use of key rate durations in measuring the sensitivity of bonds to changes in the shape of the benchmark yield curve;
☐	**e.** explain how a bond's maturity, coupon, and yield level affect its interest rate risk;
☐	**f.** calculate the duration of a portfolio and explain the limitations of portfolio duration;
☐	**g.** calculate and interpret the money duration of a bond and price value of a basis point (PVBP);
☐	**h.** calculate and interpret approximate convexity and compare approximate and effective convexity;
☐	**i.** calculate the percentage price change of a bond for a specified change in yield, given the bond's approximate duration and convexity;
☐	**j.** describe how the term structure of yield volatility affects the interest rate risk of a bond;

(continued)

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**k.** describe the relationships among a bond's holding period return, its duration, and the investment horizon;
☐	**l.** explain how changes in credit spread and liquidity affect yield-to-maturity of a bond and how duration and convexity can be used to estimate the price effect of the changes.
☐	**m.** describe the difference between empirical duration and analytical duration.

1 INTRODUCTION

Successful analysts must develop a solid understanding of the risk and return characteristics of fixed-income investments. Beyond the vast global market for public and private fixed-rate bonds, many financial assets and liabilities with known future cash flows you will encounter throughout your career are evaluated using similar principles. This analysis starts with the yield-to-maturity, or internal rate of return on future cash flows, introduced in the fixed-income valuation reading. Fixed-rate bond returns are affected by many factors, the most important of which is the full receipt of all interest and principal payments on scheduled dates. Assuming no default, return is also affected by interest rate changes that affect coupon reinvestment and the bond price if it is sold prior to maturity. Price change measures may be derived from the mathematical relationship used to calculate a bond's price. Specifically, duration estimates the price change for a given change in interest rates, and convexity improves on the duration estimate by considering that the price and yield-to-maturity relationship of a fixed-rate bond is non-linear.

Sources of return on a fixed-rate bond investment include the receipt and reinvestment of coupon payments and either the redemption of principal if the bond is held to maturity or capital gains (or losses) if the bond is sold earlier. Fixed-income investors holding the same bond may have different interest rate risk exposures if their investment horizons differ.

We introduce bond duration and convexity, showing how these statistics are calculated and used as interest rate risk measures. Although procedures and formulas exist to calculate duration and convexity, these statistics can be approximated using basic bond-pricing techniques and a financial calculator. Commonly used versions of the statistics are covered, including Macaulay, modified, effective, and key rate durations, and we distinguish between risk measures based on changes in the bond's yield-to-maturity (i.e., *yield* duration and convexity) and on benchmark yield curve changes (i.e., *curve* duration and convexity).

We then return to the investment time horizon. When an investor has a short-term horizon, duration and convexity are used to estimate the change in the bond price. Note that yield volatility matters, because bonds with varying times-to-maturity have different degrees of yield volatility. When an investor has a long-term horizon, the interaction between coupon reinvestment risk and market price risk matters. The relationship among interest rate risk, bond duration, and the investment horizon is explored.

Finally, we discuss how duration and convexity may be extended to credit and liquidity risks and highlight how these factors can affect a bond's return and risk. In addition, we highlight the use of statistical methods and historical data to establish empirical as opposed to analytical duration estimates.

SOURCES OF RETURN

2

a calculate and interpret the sources of return from investing in a fixed-rate bond

Fixed-rate bond investors have three sources of return: (1) receipt of promised coupon and principal payments on the scheduled dates, (2) reinvestment of coupon payments, and (3) potential capital gains or losses on the sale of the bond prior to maturity. In this section, it is assumed that the issuer makes the coupon and principal payments as scheduled. Here, the focus is primarily on how interest rate changes affect the reinvestment of coupon payments and a bond's market price if sold prior to maturity. Credit risk is considered later and is also the primary subject of a subsequent reading.

When a bond is purchased at a premium or a discount, it adds another aspect to the rate of return. Recall from the fixed-income valuation reading that a discount bond offers the investor a "deficient" coupon rate below the market discount rate. The amortization of this discount in each period brings the return in line with the market discount rate as the bond's carrying value is "pulled to par." For a premium bond, the coupon rate exceeds the market discount rate and the amortization of the premium adjusts the return to match the market discount rate. Through amortization, the bond's carrying value reaches par value at maturity.

A series of examples will demonstrate the effect of a change in interest rates on two investors' realized rate of returns. Interest rates are the rates at which coupon payments are reinvested and the market discount rates at the time of purchase and at the time of sale if the bond is not held to maturity. In Examples 1 and 2, interest rates are unchanged. The two investors, however, have different time horizons for holding the bond. Examples 3 and 4 show the impact of higher interest rates on the two investors' total return. Examples 5 and 6 show the impact of lower interest rates. In each of the six examples, an investor initially buys a 10-year, 8% annual coupon payment bond at a price of 85.503075 per 100 of par value. The bond's yield-to-maturity is 10.40%.

$$85.503075 = \frac{8}{(1+r)^1} + \frac{8}{(1+r)^2} + \frac{8}{(1+r)^3} + \frac{8}{(1+r)^4} + \frac{8}{(1+r)^5} +$$

$$\frac{8}{(1+r)^6} + \frac{8}{(1+r)^7} + \frac{8}{(1+r)^8} + \frac{8}{(1+r)^9} + \frac{108}{(1+r)^{10}},$$

$$r = 0.1040$$

EXAMPLE 1

A "buy-and-hold" investor purchases a 10-year, 8% annual coupon payment bond at 85.503075 per 100 of par value and holds it until maturity. The investor receives the series of 10 coupon payments of 8 (per 100 of par value) for a total of 80, plus the redemption of principal (100) at maturity. In addition to collecting the coupon interest and the principal, the investor may reinvest the cash flows. If the coupon payments are reinvested at 10.40%, the future value of the coupons on the bond's maturity date is 129.970678 per 100 of par value.

$$\left[8 \times (1.1040)^9\right] + \left[8 \times (1.1040)^8\right] + \left[8 \times (1.1040)^7\right] + \left[8 \times (1.1040)^6\right] +$$
$$\left[8 \times (1.1040)^5\right] + \left[8 \times (1.1040)^4\right] + \left[8 \times (1.1040)^3\right] + \left[8 \times (1.1040)^2\right] +$$
$$\left[8 \times (1.1040)^1\right] + 8 = 129.970678$$

The first coupon payment of 8 is reinvested at 10.40% for nine years until maturity, the second is reinvested for eight years, and so forth. The future value of the annuity is obtained easily on a financial calculator, using 8 for the payment that is received at the end of each of the 10 periods. The amount in excess of the coupons, 49.970678 (= 129.970678 – 80), is the "interest-on-interest" gain from compounding.

The investor's total return is 229.970678, the sum of the reinvested coupons (129.970678) and the redemption of principal at maturity (100). The realized rate of return is 10.40%.

$$85.503075 = \frac{229.970678}{(1 + r)^{10}}, r = 0.1040$$

Example 1 demonstrates that the yield-to-maturity at the time of purchase measures the investor's rate of return under three assumptions: (1) The investor holds the bond to maturity, (2) there is no default by the issuer, and (3) the coupon interest payments are reinvested at that same rate of interest.

Example 2 considers another investor who buys the 10-year, 8% annual coupon payment bond and pays the same price. This investor, however, has a four-year investment horizon. Therefore, coupon interest is only reinvested for four years, and the bond is sold immediately after receiving the fourth coupon payment.

EXAMPLE 2

A second investor buys the 10-year, 8% annual coupon payment bond and sells the bond after four years. Assuming that the coupon payments are reinvested at 10.40% for four years, the future value of the reinvested coupons is 37.347111 per 100 of par value.

$$\left[8 \times (1.1040)^3\right] + \left[8 \times (1.1040)^2\right] + \left[8 \times (1.1040)^1\right] + 8 = 37.347111$$

The interest-on-interest gain from compounding is 5.347111 (= 37.347111 – 32). After four years, when the bond is sold, it has six years remaining until maturity. If the yield-to-maturity remains 10.40%, the sale price of the bond is 89.668770.

$$\frac{8}{(1.1040)^1} + \frac{8}{(1.1040)^2} + \frac{8}{(1.1040)^3} + \frac{8}{(1.1040)^4} +$$
$$\frac{8}{(1.1040)^5} + \frac{8}{(1.1040)^6} = 89.668770$$

The total return is 127.015881 (= 37.347111 + 89.668770), and the realized rate of return is 10.40%.

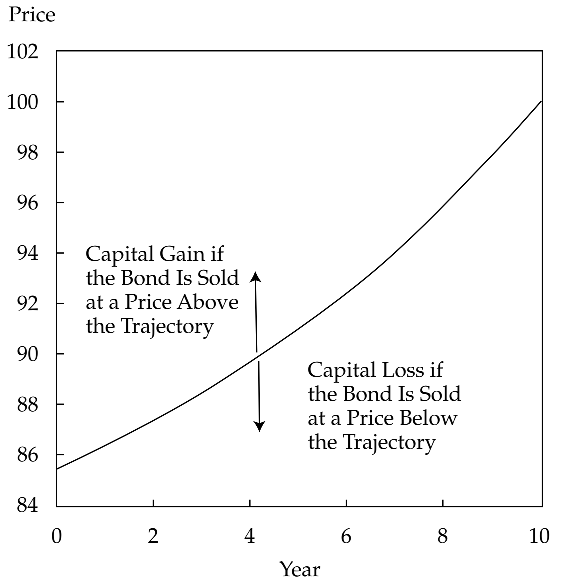
$$85.503075 = \frac{127.015881}{(1+r)^4},\ r = 0.1040$$

In Example 2, the investor's **horizon yield** is 10.40%. A horizon yield is the internal rate of return between the total return (the sum of reinvested coupon payments and the sale price or redemption amount) and the purchase price of the bond. The horizon yield on a bond investment is the annualized holding-period rate of return.

Example 2 demonstrates that the realized horizon yield matches the original yield-to-maturity if: (1) coupon payments are reinvested at the same interest rate as the original yield-to-maturity, and (2) the bond is sold at a price on the constant-yield price trajectory, which implies that the investor does not have any capital gains or losses when the bond is sold.

Capital gains arise if a bond is sold at a price above its constant-yield price trajectory and capital losses occur if a bond is sold at a price below its constant-yield price trajectory. This trajectory is based on the yield-to-maturity when the bond is purchased. The trajectory is shown in Exhibit 1 for a 10-year, 8% annual payment bond purchased at a price of 85.503075 per 100 of par value.

Exhibit 1 Constant-Yield Price Trajectory for a 10-Year, 8% Annual Payment Bond



Note: Price is price per 100 of par value.

A point on the trajectory represents the **carrying value** of the bond at that time. The carrying value is the purchase price plus the amortized amount of the discount if the bond is purchased at a price below par value. If the bond is purchased at a price above par value, the carrying value is the purchase price minus the amortized amount of the premium.

The amortized amount for each year is the change in the price between two points on the trajectory. The initial price of the bond is 85.503075 per 100 of par value. Its price (the carrying value) after one year is 86.395394, calculated using the original yield-to-maturity of 10.40%. Therefore, the amortized amount for the first year is 0.892320 (= 86.395394 – 85.503075). The bond price in Example 2 increases from 85.503075 to 89.668770, and that increase over the four years is movement *along* the constant-yield price trajectory. At the time the bond is sold, its carrying value is also 89.668770, so there is no capital gain or loss.

Examples 3 and 4 demonstrate the impact on investors' realized horizon yields if interest rates go up by 100 basis points (bps). The market discount rate on the bond increases from 10.40% to 11.40%. Coupon reinvestment rates go up by 100 bps as well.

EXAMPLE 3

The buy-and-hold investor purchases the 10-year, 8% annual payment bond at 85.503075. After the bond is purchased and before the first coupon is received, interest rates go up to 11.40%. The future value of the reinvested coupons at 11.40% for 10 years is 136.380195 per 100 of par value.

$$\left[8 \times (1.1140)^9\right] + \left[8 \times (1.1140)^8\right] + \left[8 \times (1.1140)^7\right] + \left[8 \times (1.1140)^6\right] + \left[8 \times (1.1140)^5\right] + \left[8 \times (1.1140)^4\right] + \left[8 \times (1.1140)^3\right] + \left[8 \times (1.1140)^2\right] + \left[8 \times (1.1140)^1\right] + 8 = 136.380195$$

The total return is 236.380195 (= 136.380195 + 100). The investor's realized rate of return is 10.70%.

$$85.503075 = \frac{236.380195}{(1+r)^{10}},\ r = 0.1070$$

In Example 3, the buy-and-hold investor benefits from the higher coupon reinvestment rate. The realized horizon yield is 10.70%, 30 bps higher than the outcome in Example 1, when interest rates are unchanged. There is no capital gain or loss because the bond is held until maturity. The carrying value at the maturity date is par value, the same as the redemption amount.

EXAMPLE 4

The second investor buys the 10-year, 8% annual payment bond at 85.503075 and sells it in four years. After the bond is purchased, interest rates go up to 11.40%. The future value of the reinvested coupons at 11.40% after four years is 37.899724 per 100 of par value.

$$\left[8 \times (1.1140)^3\right] + \left[8 \times (1.1140)^2\right] + \left[8 \times (1.1140)^1\right] + 8 = 37.899724$$

The sale price of the bond after four years is 85.780408.

$$\frac{8}{(1.1140)^1} + \frac{8}{(1.1140)^2} + \frac{8}{(1.1140)^3} + \frac{8}{(1.1140)^4} + \frac{8}{(1.1140)^5} + \frac{108}{(1.1140)^6} = 85.780408$$

The total return is 123.680132 (= 37.899724 + 85.780408), resulting in a realized four-year horizon yield of 9.67%.

$$85.503075 = \frac{123.680132}{(1+r)^4}, r = 0.0967$$

In Example 4, the second investor has a lower realized rate of return compared with the investor in Example 2, in which interest rates are unchanged. The future value of reinvested coupon payments goes up by 0.552613 (= 37.899724 – 37.347111) per 100 of par value because of the higher interest rates. But there is a *capital loss* of 3.888362 (= 89.668770 – 85.780408) per 100 of par value. Notice that the capital loss is measured from the bond's carrying value, the point on the constant-yield price trajectory, and not from the original purchase price. The bond is now sold at a price below the constant-yield price trajectory. The reduction in the realized four-year horizon yield from 10.40% to 9.67% is a result of the capital loss being greater than the gain from reinvesting coupons at a higher rate, which reduces the investor's total return.

Examples 5 and 6 complete the series of rate-of-return calculations for the two investors. Interest rates decline by 100 bps. The required yield on the bond falls from 10.40% to 9.40% after the purchase of the bond. The interest rates at which the coupon payments are reinvested fall as well.

EXAMPLE 5

The buy-and-hold investor purchases the 10-year bond at 85.503075 and holds the security until it matures. After the bond is purchased and before the first coupon is received, interest rates go down to 9.40%. The future value of reinvesting the coupon payments at 9.40% for 10 years is 123.888356 per 100 of par value.

$$\left[8 \times (1.0940)^9\right] + \left[8 \times (1.0940)^8\right] + \left[8 \times (1.0940)^7\right] + \left[8 \times (1.0940)^6\right] +$$
$$\left[8 \times (1.0940)^5\right] + \left[8 \times (1.0940)^4\right] + \left[8 \times (1.0940)^3\right] + \left[8 \times (1.0940)^2\right] +$$
$$\left[8 \times (1.0940)^1\right] + 8 = 123.888356$$

The total return is 223.888356, the sum of the future value of reinvested coupons and the redemption of par value. The investor's realized rate of return is 10.10%.

$$85.503075 = \frac{223.888365}{(1+r)^{10}}, r = 0.1010$$

In Example 5, the buy-and-hold investor suffers from the lower coupon reinvestment rates. The realized horizon yield is 10.10%, 30 bps lower than the result in Example 1, when interest rates are unchanged. There is no capital gain or loss because the bond is held until maturity. Examples 1, 3, and 5 indicate that the interest rate risk for a buy-and-hold investor arises entirely from changes in coupon reinvestment rates.

EXAMPLE 6

The second investor buys the 10-year bond at 85.503075 and sells it in four years. After the bond is purchased, interest rates go down to 9.40%. The future value of the reinvested coupons at 9.40% is 36.801397 per 100 of par value.

$$\left[8\times(1.0940)^3\right]+\left[8\times(1.0940)^2\right]+\left[8\times(1.0940)^1\right]+8=36.801397$$

This reduction in future value is offset by the higher sale price of the bond, which is 93.793912 per 100 of par value.

$$\frac{8}{(1.0940)^1}+\frac{8}{(1.0940)^2}+\frac{8}{(1.0940)^3}+\frac{8}{(1.0940)^4}+$$
$$\frac{8}{(1.0940)^5}+\frac{108}{(1.0940)^6}=93.793912$$

The total return is 130.595309 (= 36.801397 + 93.793912), and the realized yield is 11.17%.

$$85.503075=\frac{130.595309}{(1+r)^4},\quad r=0.1117$$

The investor in Example 6 has a capital gain of 4.125142 (= 93.793912 – 89.668770). The capital gain is measured from the carrying value, the point on the constant-yield price trajectory. That gain offsets the reduction in the future value of reinvested coupons of 0.545714 (= 37.347111 – 36.801397). The total return is higher than that in Example 2, in which the interest rate remains at 10.40%.

In these examples, interest income for the investor is the return associated with the *passage of time*. Therefore, interest income includes the receipt of coupon interest, the reinvestment of those cash flows, and the amortization of the discount from purchase at a price below par value (or the premium from purchase at a price above par value) to bring the return back in line with the market discount rate. A capital gain or loss is the return to the investor associated with the *change in the value* of the security. On the fixed-rate bond, a change in value arises from a change in the yield-to-maturity, which is the implied market discount rate. In practice, the way interest income and capital gains and losses are calculated and reported on financial statements depends on financial and tax accounting rules.

This series of examples illustrates an important point about fixed-rate bonds: The *investment horizon* is at the heart of understanding interest rate risk and return. There are two offsetting types of interest rate risk that affect the bond investor: coupon reinvestment risk and market price risk. The future value of reinvested coupon payments (and, in a portfolio, the principal on bonds that mature before the horizon date) *increases* when interest rates rise and *decreases* when rates fall. The sale price on a bond that matures after the horizon date (and thus needs to be sold) *decreases* when interest rates rise and *increases* when rates fall. Coupon reinvestment risk matters more when the investor has a long-term horizon relative to the time-to-maturity of the bond. For instance, a buy-and-hold investor only has coupon reinvestment risk. Market price risk matters more when the investor has a short-term horizon relative to the time-to-maturity. For example, an investor who sells the bond before the first coupon is received has only market price risk. Therefore, two investors holding the same bond (or bond portfolio) can have different exposures to interest rate risk if they have different investment horizons.

EXAMPLE 7

An investor buys a four-year, 10% annual coupon payment bond priced to yield 5.00%. The investor plans to sell the bond in two years once the second coupon payment is received. Calculate the purchase price for the bond and the horizon yield assuming that the coupon reinvestment rate after the bond purchase and the yield-to-maturity at the time of sale are (1) 3.00%, (2) 5.00%, and (3) 7.00%.

Solution:

The purchase price is 117.729753.

$$\frac{10}{(1.0500)^1} + \frac{10}{(1.0500)^2} + \frac{10}{(1.0500)^3} + \frac{110}{(1.0500)^4} = 117.729753$$

1 3.00%: The future value of reinvested coupons is 20.300.

$$(10 \times 1.0300) + 10 = 20.300$$

The sale price of the bond is 113.394288.

$$\frac{10}{(1.0300)^1} + \frac{110}{(1.0300)^2} = 113.394288$$

Total return: 20.300 + 113.394288 = 133.694288.

If interest rates go down from 5.00% to 3.00%, the realized rate of return over the two-year investment horizon is 6.5647%, higher than the original yield-to-maturity of 5.00%.

$$117.729753 = \frac{133.694288}{(1 + r)^2}, \quad r = 0.065647$$

2 5.00%: The future value of reinvested coupons is 20.500.

$$(10 \times 1.0500) + 10 = 20.500$$

The sale price of the bond is 109.297052.

$$\frac{10}{(1.0500)^1} + \frac{110}{(1.0500)^2} = 109.297052$$

Total return: 20.500 + 109.297052 = 129.797052.

If interest rates remain 5.00% for reinvested coupons and for the required yield on the bond, the realized rate of return over the two-year investment horizon is equal to the yield-to-maturity of 5.00%.

$$117.729753 = \frac{129.797052}{(1 + r)^2}, \quad r = 0.050000$$

3 7.00%: The future value of reinvested coupons is 20.700.

$$(10 \times 1.0700) + 10 = 20.700$$

The bond is sold at 105.424055.

$$\frac{10}{(1.0700)^1} + \frac{110}{(1.0700)^2} = 105.424055$$

Total return: 20.700 + 105.424055 = 126.124055.

$$117.729753 = \frac{126.124055}{(1+r)^2}, \quad r = 0.035037$$

If interest rates go up from 5.00% to 7.00%, the realized rate of return over the two-year investment horizon is 3.5037%, lower than the yield-to-maturity of 5.00%.

3 MACAULAY AND MODIFIED DURATION

b define, calculate, and interpret Macaulay, modified, and effective durations

This section covers two commonly used measures of interest rate risk: duration and convexity. It distinguishes between risk measures based on changes in a bond's own yield-to-maturity (yield duration and convexity) and those that affect the bond based on changes in a benchmark yield curve (curve duration and convexity).

3.1 Macaulay, Modified, and Approximate Duration

The duration of a bond measures the sensitivity of the bond's full price (including accrued interest) to changes in the bond's yield-to-maturity or, more generally, to changes in benchmark interest rates. Duration estimates changes in the bond price assuming that variables other than the yield-to-maturity or benchmark rates are held constant. Most importantly, the time-to-maturity is unchanged. Therefore, duration measures the *instantaneous* (or, at least, same-day) change in the bond price. The accrued interest is the same, so it is the flat price that goes up or down when the full price changes. Duration is a useful measure because it represents the approximate amount of time a bond would have to be held for the market discount rate at purchase to be realized if there is a single change in interest rate. If the bond is held for the duration period, an increase from reinvesting coupons is offset by a decrease in price if interest rates increase and a decrease from reinvesting coupons is offset by an increase in price if interest rates decrease.

There are several types of bond duration. In general, these can be divided into **yield duration** and **curve duration**. Yield duration is the sensitivity of the bond price with respect to the bond's own yield-to-maturity. Curve duration is the sensitivity of the bond price (or more generally, the market value of a financial asset or liability) with respect to a benchmark yield curve. The benchmark yield curve could be the government yield curve on coupon bonds, the spot curve, or the forward curve, but in practice, the government par curve is often used. Yield duration statistics used in fixed-income analysis include Macaulay duration, modified duration, money duration, and the price value of a basis point (PVBP). A curve duration statistic often used is effective duration. Effective duration is covered later in this reading.

Macaulay duration is named after Frederick Macaulay, the Canadian economist who first wrote about the statistic in 1938. Equation 1 is a general formula to calculate the Macaulay duration (MacDur) of a traditional fixed-rate bond.

$$\text{MacDur} = \left[\frac{\dfrac{(1-t/T)\times PMT}{(1+r)^{1-t/T}} + \dfrac{(2-t/T)\times PMT}{(1+r)^{2-t/T}} + \cdots + \dfrac{(N-t/T)\times(PMT+FV)}{(1+r)^{N-t/T}}}{\dfrac{PMT}{(1+r)^{1-t/T}} + \dfrac{PMT}{(1+r)^{2-t/T}} + \cdots + \dfrac{PMT+FV}{(1+r)^{N-t/T}}}\right] \quad \textbf{(1)}$$

where

t = the number of days from the last coupon payment to the settlement date

T = the number of days in the coupon period

t/T = the fraction of the coupon period that has gone by since the last payment

PMT = the coupon payment per period

FV = the future value paid at maturity, or the par value of the bond

r = the yield-to-maturity, or the market discount rate, per period

N = the number of evenly spaced periods to maturity as of the beginning of the current period

The denominator in Equation 1 is the full price (PV^{Full}) of the bond including accrued interest. It is the present value of the coupon interest and principal payments, with each cash flow discounted by the same market discount rate, r.

$$PV^{Full} = \frac{PMT}{(1+r)^{1-t/T}} + \frac{PMT}{(1+r)^{2-t/T}} + \cdots + \frac{PMT+FV}{(1+r)^{N-t/T}} \quad \textbf{(2)}$$

Equation 3 combines Equations 1 and 2 to reveal an important aspect of the Macaulay duration: Macaulay duration is a weighted average of the time to receipt of the bond's promised payments, where the weights are the shares of the full price that correspond to each of the bond's promised future payments.

$$\text{MacDur} = \left\{\begin{array}{l} (1-t/T)\left[\dfrac{\dfrac{PMT}{(1+r)^{1-t/T}}}{PV^{Full}}\right] + (2-t/T)\left[\dfrac{\dfrac{PMT}{(1+r)^{2-t/T}}}{PV^{Full}}\right] + \cdots + \\ (N-t/T)\left[\dfrac{\dfrac{PMT+FV}{(1+r)^{N-t/T}}}{PV^{Full}}\right] \end{array}\right\} \quad \textbf{(3)}$$

The times to receipt of cash flow measured in terms of time periods are $1 - t/T$, $2 - t/T, \ldots, N - t/T$. The weights are the present values of the cash flows divided by the full price. Therefore, Macaulay duration is measured in terms of time periods. A couple of examples will clarify this calculation.

Consider first the 10-year, 8% annual coupon payment bond used in Examples 1–6. The bond's yield-to-maturity is 10.40%, and its price is 85.503075 per 100 of par value. This bond has 10 evenly spaced periods to maturity. Settlement is on a coupon payment date so that $t/T = 0$. Exhibit 2 illustrates the calculation of the bond's Macaulay duration.

Exhibit 2 Macaulay Duration of a 10-Year, 8% Annual Payment Bond

Period	Cash Flow	Present Value	Weight	Period × Weight
1	8	7.246377	0.08475	0.0847
2	8	6.563747	0.07677	0.1535
3	8	5.945423	0.06953	0.2086
4	8	5.385347	0.06298	0.2519
5	8	4.878032	0.05705	0.2853
6	8	4.418507	0.05168	0.3101
7	8	4.002271	0.04681	0.3277
8	8	3.625245	0.04240	0.3392
9	8	3.283737	0.03840	0.3456
10	108	40.154389	0.46963	4.6963
		85.503075	1.00000	7.0029

The first two columns of Exhibit 2 show the number of periods to the receipt of the cash flow and the amount of the payment per 100 of par value. The third column is the present value of the cash flow. For example, the final payment is 108 (the last coupon payment plus the redemption of principal) and its present value is 40.154389.

$$\frac{108}{(1.1040)^{10}} = 40.154389$$

The sum of the present values is the full price of the bond. The fourth column is the weight, the share of total market value corresponding to each cash flow. The final payment of 108 per 100 of par value is 46.963% of the bond's market value.

$$\frac{40.154389}{85.503075} = 0.46963$$

The sum of the weights is 1.00000. The fifth column is the number of periods to the receipt of the cash flow (the first column) multiplied by the weight (the fourth column). The sum of that column is 7.0029, which is the Macaulay duration of this 10-year, 8% annual coupon payment bond. This statistic is sometimes reported as 7.0029 *years,* although the time frame is not needed in most applications.

Now consider an example *between* coupon payment dates. A 6% semiannual payment corporate bond that matures on 14 February 2027 is purchased for settlement on 11 April 2019. The coupon payments are 3 per 100 of par value, paid on 14 February and 14 August of each year. The yield-to-maturity is 6.00% quoted on a street-convention semiannual bond basis. The full price of this bond comprises the flat price plus accrued interest. The flat price for the bond is 99.990423 per 100 of par value. The accrued interest is calculated using the 30/360 method to count days. This settlement date is 57 days into the 180-day semiannual period, so t/T = 57/180. The accrued interest is 0.950000 (= 57/180 × 3) per 100 of par value. The full price for the bond is 100.940423 (= 99.990423 + 0.950000). Exhibit 3 shows the calculation of the bond's Macaulay duration.

Exhibit 3 Macaulay Duration of an Eight-Year, 6% Semiannual Payment Bond Priced to Yield 6.00%

Period	Time to Receipt	Cash Flow	Present Value	Weight	Time × Weight
1	0.6833	3	2.940012	0.02913	0.019903
2	1.6833	3	2.854381	0.02828	0.047601
3	2.6833	3	2.771244	0.02745	0.073669
4	3.6833	3	2.690528	0.02665	0.098178
5	4.6833	3	2.612163	0.02588	0.121197
6	5.6833	3	2.536080	0.02512	0.142791
7	6.6833	3	2.462214	0.02439	0.163025
8	7.6833	3	2.390499	0.02368	0.181959
9	8.6833	3	2.320873	0.02299	0.199652
10	9.6833	3	2.253275	0.02232	0.216159
11	10.6833	3	2.187645	0.02167	0.231536
12	11.6833	3	2.123927	0.02104	0.245834
13	12.6833	3	2.062065	0.02043	0.259102
14	13.6833	3	2.002005	0.01983	0.271389
15	14.6833	3	1.943694	0.01926	0.282740
16	15.6833	103	64.789817	0.64186	10.066535
			100.940423	1.00000	12.621268

There are 16 semiannual periods to maturity between the last coupon payment date of 14 February 2019 and maturity on 14 February 2027. The time to receipt of cash flow in semiannual periods is in the second column: 0.6833 = 1 – 57/180, 1.6833 = 2 – 57/180, etc. The cash flow for each period is in the third column. The annual yield-to-maturity is 6.00%, so the yield per semiannual period is 3.00%. When that yield is used to get the present value of each cash flow, the full price of the bond is 100.940423, the sum of the fourth column. The weights, which are the shares of the full price corresponding to each cash flow, are in the fifth column. The Macaulay duration is the sum of the items in the sixth column, which is the weight multiplied by the time to receipt of each cash flow. The result, 12.621268, is the Macaulay duration on an eight-year, 6% semiannual payment bond for settlement on 11 April 2019 measured in *semiannual periods*. Similar to coupon rates and yields-to-maturity, duration statistics invariably are annualized in practice. Therefore, the Macaulay duration typically is reported as 6.310634 *years* (= 12.621268/2). (Such precision for the duration statistic is not needed in practice. Typically, "6.31 years" is enough. The full precision is shown here to illustrate calculations.) Microsoft Excel users can obtain the Macaulay duration using the DURATION financial function—DURATION(DATE(2019,4,11),DATE(2027,2,14),0.06,0.06,2,0)—and inputs that include the settlement date, maturity date, annual coupon rate as a decimal, annual yield-to-maturity as a decimal, periodicity, and day count code (0 for 30/360, 1 for actual/actual).

Another approach to calculating the Macaulay duration is to use a closed-form equation derived using calculus and algebra (see Smith 2014). Equation 4 is a general closed-form formula for determining the Macaulay duration of a fixed-rate bond, where c is the coupon rate per period (PMT/FV).

$$\text{MacDur} = \left\{\frac{1+r}{r} - \frac{1+r+\left[N\times(c-r)\right]}{c\times\left[(1+r)^N-1\right]+r}\right\} - (t/T) \tag{4}$$

The Macaulay duration of the 10-year, 8% annual payment bond is calculated by entering r = 0.1040, c = 0.0800, N = 10, and t/T = 0 into Equation 4.

$$\text{MacDur} = \frac{1+0.1040}{0.1040} - \frac{1+0.1040+\left[10\times(0.0800-0.1040)\right]}{0.0800\times\left[(1+0.1040)^{10}-1\right]+0.1040} = 7.0029$$

Therefore, the weighted average time to receipt of the interest and principal payments that will result in realization of the initial market discount rate on this 10-year bond is 7.00 years.

The Macaulay duration of the 6% semiannual payment bond maturing on 14 February 2027 is obtained by entering r = 0.0300, c = 0.0300, N = 16, and t/T = 57/180 into Equation 4.

$$\text{MacDur} = \left[\frac{1+0.0300}{0.0300} - \frac{1+0.0300+\left[16\times(0.0300-0.0300)\right]}{0.0300\times\left[(1+0.0300)^{16}-1\right]+0.0300}\right] - (57/180)$$
$$= 12.621268$$

Equation 4 uses the yield-to-maturity *per period*, the coupon rate *per period*, the number of *periods* to maturity, and the fraction of the current *period* that has gone by. Its output is the Macaulay duration in terms of *periods*. It is converted to annual duration by dividing by the number of periods in the year.

The calculation of the **modified duration** (ModDur) statistic of a bond requires a simple adjustment to Macaulay duration. It is the Macaulay duration statistic divided by one plus the yield per period.

$$\text{ModDur} = \frac{\text{MacDur}}{1+r} \tag{5}$$

For example, the modified duration of the 10-year, 8% annual payment bond is 6.3432.

$$\text{ModDur} = \frac{7.0029}{1.1040} = 6.3432$$

The modified duration of the 6% semiannual payment bond maturing on 14 February 2027 is 12.253658 semiannual periods.

$$\text{ModDur} = \frac{12.621268}{1.0300} = 12.253658$$

The annualized modified duration of the bond is 6.126829 (= 12.253658/2).

Microsoft Excel users can obtain the modified duration using the MDURATION financial function using the same inputs as for the Macaulay duration: MDURATION(DATE(2019,4,11),DATE(2027,2,14),0.06,0.06,2,0). Although modified duration might seem to be just a Macaulay duration with minor adjustments, it has an important application in risk measurement: Modified duration provides an estimate of the percentage price change for a bond given a change in its yield-to-maturity.

$$\%\Delta PV^{Full} \approx -\text{AnnModDur} \times \Delta\text{Yield} \tag{6}$$

The percentage price change refers to the full price, including accrued interest. The AnnModDur term in Equation 6 is the *annual* modified duration, and the ΔYield term is the change in the *annual* yield-to-maturity. The ≈ sign indicates that this calculation is an estimation. The minus sign indicates that bond prices and yields-to-maturity move inversely.

If the annual yield on the 6% semiannual payment bond that matures on 14 February 2027 jumps by 100 bps, from 6.00% to 7.00%, the estimated loss in value for the bond is 6.1268%.

$$\%\Delta PV^{Full} \approx -6.126829 \times 0.0100 = -0.061268$$

If the yield-to-maturity were to drop by 100 bps to 5.00%, the estimated gain in value is also 6.1268%.

$$\%\Delta PV^{Full} \approx -6.126829 \times -0.0100 = 0.061268$$

Modified duration provides a *linear* estimate of the percentage price change. In terms of absolute value, the change is the same for either an increase or a decrease in the yield-to-maturity. Recall that for a given coupon rate and time-to-maturity, the percentage price change is greater (in absolute value) when the market discount rate goes down than when it goes up. Later in this reading, a "convexity adjustment" to duration is introduced. It improves the accuracy of this estimate, especially when a large change in yield-to-maturity (such as 100 bps) is considered.

APPROXIMATE MODIFIED AND MACAULAY DURATION

4

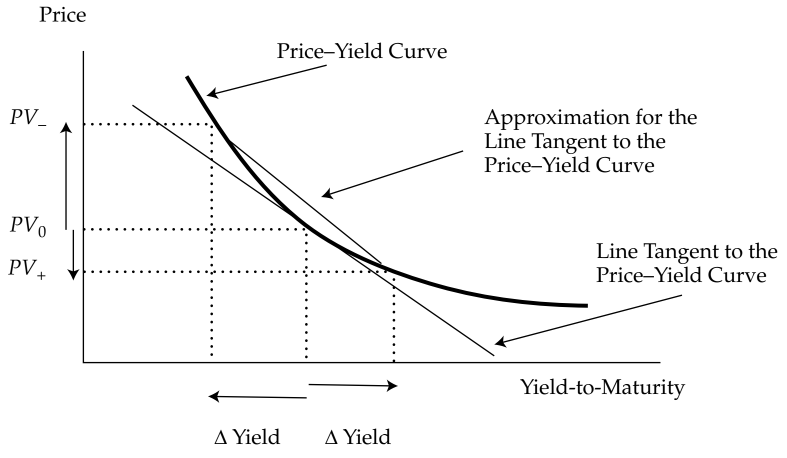
b define, calculate, and interpret Macaulay, modified, and effective durations

The modified duration statistic for a fixed-rate bond is easily obtained if the Macaulay duration is already known. An alternative approach is to *approximate* modified duration directly. Equation 7 is the approximation formula for annual modified duration.

$$\text{ApproxModDur} = \frac{(PV_-) - (PV_+)}{2 \times (\Delta\text{Yield}) \times (PV_0)} \quad (7)$$

The objective of the approximation is to estimate the slope of the line tangent to the price–yield curve. The slope of the tangent and the approximated slope are shown in Exhibit 4.

Exhibit 4 Approximate Modified Duration



To estimate the slope, the yield-to-maturity is changed up and down by the same amount—the ΔYield. Then the bond prices given the new yields-to-maturity are calculated. The price when the yield is increased is denoted PV_+. The price when the yield-to-maturity is reduced is denoted PV_-. The original price is PV_0. These prices are the full prices, including accrued interest. The slope of the line based on PV_+ and PV_- is the approximation for the slope of the line tangent to the price–yield curve. The following example illustrates the remarkable accuracy of this approximation. In fact, as ΔYield approaches zero, the approximation approaches AnnModDur.

Consider the 6% semiannual coupon payment corporate bond maturing on 14 February 2027. For settlement on 11 April 2019, the full price (PV_0) is 100.940423 given that the yield-to-maturity is 6.00%.

$$PV_0 = \left[\frac{3}{(1.03)^1} + \frac{3}{(1.03)^2} + \cdots + \frac{103}{(1.03)^{16}}\right] \times (1.03)^{57/180} = 100.940423$$

Raise the annual yield-to-maturity by five bps, from 6.00% to 6.05%. This increase corresponds to an increase in the yield-to-maturity per semiannual period of 2.5 bps, from 3.00% to 3.025% per period. The new full price (PV_+) is 100.631781.

$$PV_+ = \left[\frac{3}{(1.03025)^1} + \frac{3}{(1.03025)^2} + \cdots + \frac{103}{(1.03025)^{16}}\right] \times (1.03025)^{57/180} = 100.631781$$

Lower the annual yield-to-maturity by five bps, from 6.00% to 5.95%. This decrease corresponds to a decrease in the yield-to-maturity per semiannual period of 2.5 bps, from 3.00% to 2.975% per period. The new full price (PV_-) is 101.250227.

$$PV_- = \left[\frac{3}{(1.02975)^1} + \frac{3}{(1.02975)^2} + \cdots + \frac{103}{(1.02975)^{16}}\right] \times (1.02975)^{57/180} = 101.250227$$

Enter these results into Equation 7 for the 5 bp change in the annual yield-to-maturity, or ΔYield = 0.0005:

$$\text{ApproxModDur} = \frac{101.250227 - 100.631781}{2 \times 0.0005 \times 100.940423} = 6.126842$$

The "exact" annual modified duration for this bond is 6.126829 and the "approximation" is 6.126842—virtually identical results. Therefore, although duration can be calculated using the approach in Exhibits 2 and 3—basing the calculation on the weighted average time to receipt of each cash flow—or using the closed-form formula as in Equation 4, it can also be estimated quite accurately using the basic bond-pricing equation and a financial calculator. The Macaulay duration can be approximated as well—the approximate modified duration multiplied by one plus the yield per period.

$$\text{ApproxMacDur} = \text{ApproxModDur} \times (1 + r) \tag{8}$$

The approximation formulas produce results for *annualized* modified and Macaulay durations. The frequency of coupon payments and the periodicity of the yield-to-maturity are included in the bond price calculations.

EXAMPLE 8

Assume that the 3.75% US Treasury bond that matures on 15 August 2041 is priced to yield 5.14% for settlement on 15 October 2020. Coupons are paid semiannually on 15 February and 15 August. The yield-to-maturity is stated on a street-convention semiannual bond basis. This settlement date is 61 days into a 184-day coupon period, using the actual/actual day-count convention. Compute the approximate modified duration and the approximate Macaulay duration for this Treasury bond assuming a 5 bp change in the yield-to-maturity.

Solution:

The yield-to-maturity per semiannual period is 0.0257 (= 0.0514/2). The coupon payment per period is 1.875 (= 3.75/2). At the beginning of the period, there are 21 years (42 semiannual periods) to maturity. The fraction of the period that has passed is 61/184. The full price at that yield-to-maturity is 82.967530 per 100 of par value.

$$PV_0 = \left[\frac{1.875}{(1.0257)^1} + \frac{1.875}{(1.0257)^2} + \cdots + \frac{101.875}{(1.0257)^{42}}\right] \times (1.0257)^{61/184} = 82.96753$$

Raise the yield-to-maturity from 5.14% to 5.19%—therefore, from 2.57% to 2.595% per semiannual period—and the price becomes 82.411395 per 100 of par value.

$$PV_+ = \left[\frac{1.875}{(1.02595)^1} + \frac{1.875}{(1.02595)^2} + \cdots + \frac{101.875}{(1.02595)^{42}}\right] \times (1.02595)^{61/184}$$
$$= 82.411395$$

Lower the yield-to-maturity from 5.14% to 5.09%—therefore, from 2.57% to 2.545% per semiannual period—and the price becomes 83.528661 per 100 of par value.

$$PV_- = \left[\frac{1.875}{(1.02545)^1} + \frac{1.875}{(1.02545)^2} + \cdots + \frac{101.875}{(1.02545)^{42}}\right] \times (1.02545)^{61/184}$$
$$= 83.528661$$

The approximate annualized modified duration for the Treasury bond is 13.466.

$$\text{ApproxModDur} = \frac{83.528661 - 82.411395}{2 \times 0.0005 \times 82.967530} = 13.466$$

The approximate annualized Macaulay duration is 13.812.

$$\text{ApproxMacDur} = 13.466 \times 1.0257 = 13.812$$

Therefore, from these statistics, the investor knows that the weighted average time to receipt of interest and principal payments is 13.812 years (the Macaulay duration) and that the estimated loss in the bond's market value is 13.466% (the modified duration) if the market discount rate were to suddenly go up by 1% from 5.14% to 6.14%.

5 EFFECTIVE AND KEY RATE DURATION

b define, calculate, and interpret Macaulay, modified, and effective durations

c explain why effective duration is the most appropriate measure of interest rate risk for bonds with embedded options

Another approach to assess the interest rate risk of a bond is to estimate the percentage change in price given a change in a benchmark yield curve—for example, the government par curve. This estimate, which is very similar to the formula for approximate modified duration, is called the **effective duration**. The effective duration of a bond is the sensitivity of the bond's price to a change in a benchmark yield curve. The formula to calculate effective duration (EffDur) is Equation 9.

$$\text{EffDur} = \frac{(PV_{-}) - (PV_{+})}{2 \times (\Delta\text{Curve}) \times (PV_0)} \tag{9}$$

The difference between approximate modified duration and effective duration is in the denominator. Modified duration is a *yield duration* statistic in that it measures interest rate risk in terms of a change in the bond's own yield-to-maturity (ΔYield). Effective duration is a *curve duration* statistic in that it measures interest rate risk in terms of a parallel shift in the benchmark yield curve (ΔCurve).

Effective duration is essential to the measurement of the interest rate risk of a complex bond, such as a bond that contains an embedded call option. The duration of a callable bond is *not* the sensitivity of the bond price to a change in the yield-to-worst (i.e., the lowest of the yield-to-maturity, yield-to-first-call, yield-to-second-call, and so forth). The problem is that future cash flows are uncertain because they are contingent on future interest rates. The issuer's decision to call the bond depends on the ability to refinance the debt at a lower cost of funds. In brief, a callable bond does not have a well-defined internal rate of return (yield-to-maturity). Therefore, yield duration statistics, such as modified and Macaulay durations, do not apply; effective duration is the appropriate duration measure.

The specific option-pricing models that are used to produce the inputs to effective duration for a callable bond are covered in later readings. However, as an example, suppose that the full price of a callable bond is 101.060489 per 100 of par value. The option-pricing model inputs include (1) the length of the call protection period, (2) the schedule of call prices and call dates, (3) an assumption about credit spreads over benchmark yields (which includes any liquidity spread as well), (4) an assumption about future interest rate volatility, and (5) the level of market interest rates (e.g., the government par curve). The analyst then holds the first four inputs constant and raises and lowers the fifth input. Suppose that when the government par curve is raised and lowered by 25 bps, the new full prices for the callable bond from the model are 99.050120 and 102.890738, respectively. Therefore, PV_0 = 101.060489, PV_+ = 99.050120, PV_- = 102.890738, and ΔCurve = 0.0025. The effective duration for the callable bond is 7.6006.

$$\text{EffDur} = \frac{102.890738 - 99.050120}{2 \times 0.0025 \times 101.060489} = 7.6006$$

This curve duration measure indicates the bond's sensitivity to the benchmark yield curve—in particular, the government par curve—assuming no change in the credit spread. In practice, a callable bond issuer might be able to exercise the call option and obtain a lower cost of funds if (1) benchmark yields fall and the credit spread over the benchmark is unchanged or (2) benchmark yields are unchanged and the credit spread is reduced (e.g., because of an upgrade in the issuer's rating). A pricing model can be used to determine a "credit duration" statistic—that is, the sensitivity of the bond price to a change in the credit spread. On a traditional fixed-rate bond, modified duration estimates the percentage price change for a change in the benchmark yield and/or the credit spread. For bonds that do not have a well-defined internal rate of return because the future cash flows are not fixed—for instance, callable bonds and floating-rate notes—pricing models are used to produce different statistics for changes in benchmark interest rates and for changes in credit risk.

Another fixed-income security for which yield duration statistics, such as modified and Macaulay durations, are not relevant is a mortgage-backed bond. These securities arise from a residential (or commercial) loan portfolio securitization. The key point for measuring interest rate risk on a mortgage-backed bond is that the cash flows are contingent on homeowners' ability to refinance their debt at a lower rate. In effect, the homeowners have call options on their mortgage loans.

A practical consideration in using effective duration is in setting the change in the benchmark yield curve. With approximate modified duration, accuracy is improved by choosing a smaller yield-to-maturity change. But the pricing models for more-complex securities, such as callable and mortgage-backed bonds, include assumptions about the behavior of the corporate issuers, businesses, or homeowners. Rates typically need to change by a minimum amount to affect the decision to call a bond or refinance a mortgage loan because issuing new debt involves transaction costs. Therefore, estimates of interest rate risk using effective duration are not necessarily improved by choosing a smaller change in benchmark rates. Effective duration has become an important tool in the financial analysis of not only traditional bonds but also financial liabilities. Example 9 demonstrates such an application of effective duration.

EXAMPLE 9

Defined-benefit pension schemes typically pay retirees a monthly amount based on their wage level at the time of retirement. The amount could be fixed in nominal terms or indexed to inflation. These programs are referred to as "defined-benefit pension plans" when US GAAP or IFRS accounting standards are used. In Australia, they are called "superannuation funds."

A British defined-benefit pension scheme seeks to measure the sensitivity of its retirement obligations to market interest rate changes. The pension scheme manager hires an actuarial consultancy to model the present value of its liabilities under three interest rate scenarios: (1) a base rate of 5%, (2) a 100 bp increase in rates, up to 6%, and (3) a 100 bp drop in rates, down to 4%.

The actuarial consultancy uses a complex valuation model that includes assumptions about employee retention, early retirement, wage growth, mortality, and longevity. The following chart shows the results of the analysis.

Interest Rate Assumption	Present Value of Liabilities
4%	GBP973.5 million
5%	GBP926.1 million
6%	GBP871.8 million

Compute the effective duration of the pension scheme's liabilities.

Solution:

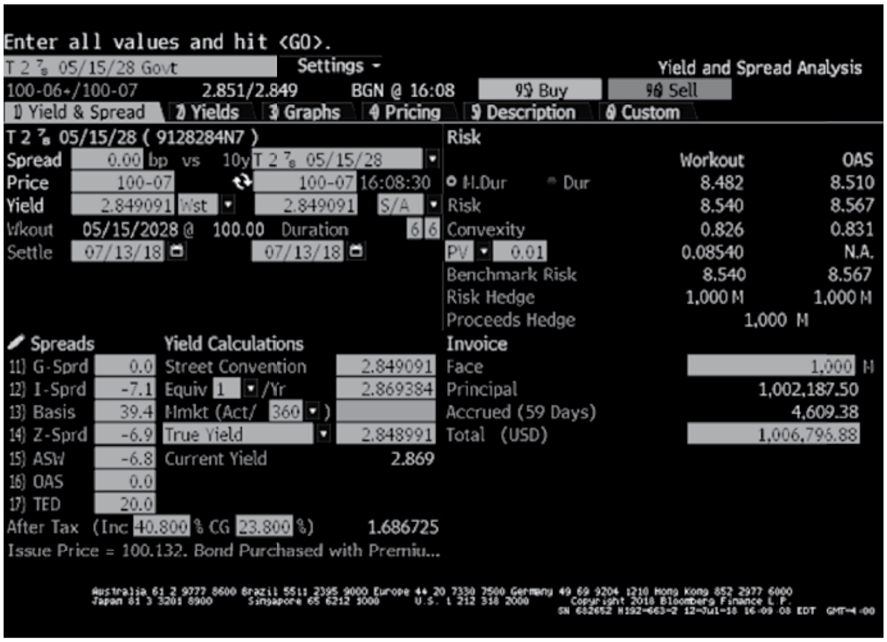
PV_0 = 926.1, PV_+ = 871.8, PV_- = 973.5, and ΔCurve = 0.0100. The effective duration of the pension scheme's liabilities is 5.49.

$$\text{EffDur} = \frac{973.5 - 871.8}{2 \times 0.0100 \times 926.1} = 5.49$$

This effective duration statistic for the pension scheme's liabilities might be used in asset allocation decisions to decide the mix of equity, fixed income, and alternative assets.

Although effective duration is the most appropriate interest rate risk measure for bonds with embedded options, it also is useful with traditional bonds to supplement the information provided by the Macaulay and modified yield durations. Exhibit 5 displays the Bloomberg Yield and Spread (YAS) Analysis page for the 2.875% US Treasury note that matures on 15 May 2028.

Exhibit 5 Bloomberg YAS Page for the 2.875% US Treasury Note



In Exhibit 5, the quoted (flat) asked price for the bond is 100-07, which is equal to 100 and 7/32nds per 100 of par value for settlement on 13 July 2018. Most bond prices are stated in decimals, but US Treasuries are usually quoted in fractions. As a decimal, the flat price is 100.21875. The accrued interest uses the actual/actual day-count method. That settlement date is 59 days into a 184-day semiannual coupon payment period. The accrued interest is 0.4609375 per 100 of par value (= 59/184 × 0.02875/2 × 100). The full price of the bond is 100.679688. The yield-to-maturity of the bond is 2.849091%, stated on a street-convention semiannual bond basis.

The modified duration for the bond is shown in Exhibit 5 to be 8.482, which is the conventional *yield* duration statistic. Its *curve* duration, however, is 8.510, which is the price sensitivity with respect to changes in the US Treasury par curve. On Bloomberg, the effective duration is called the "OAS duration" because it is based on the option-pricing model that is also used to calculate the option-adjusted spread. The small difference arises because the government yield curve is not flat. When the par curve is shifted in the model, the government spot curve is also shifted, although not in the same "parallel" manner. Therefore, the change in the bond price is not the same as it would be if its own yield-to-maturity changed by the same amount as the change in the par curve. In general, the modified duration and effective duration on a traditional option-free bond are not identical. The difference narrows when the yield curve is flatter, the time-to-maturity is shorter, and the bond is priced closer to par value (so that the difference between the coupon rate and the yield-to-maturity is smaller). The modified duration and effective duration on an option-free bond are identical only in the rare circumstance of a flat yield curve.

5.1 Key Rate Duration

d define key rate duration and describe the use of key rate durations in measuring the sensitivity of bonds to changes in the shape of the benchmark yield curve

The effective duration for a sample callable bond was calculated previously as

$$\text{EffDur} = \frac{102.890738 - 99.050120}{2 \times 0.0025 \times 101.060489} = 7.6006$$

This duration measure indicates the bond's sensitivity to the benchmark yield curve if all yields change by the same amount. "Key rate" duration provides further insight into a bond's sensitivity to non-parallel benchmark yield curve changes. A **key rate duration** (or **partial duration**) is a measure of a bond's sensitivity to a change in the benchmark yield at a specific maturity. Key rate durations define a security's price sensitivity over a set of maturities along the yield curve, with the sum of key rate durations being identical to the effective duration:

$$\text{KeyRateDur}^k = -\frac{1}{PV} \times \frac{\Delta PV}{\Delta r^k} \tag{10}$$

$$\sum_{k=1}^{n} \text{KeyRateDur}^k = \text{EffDur}$$

where r^k represents the kth key rate. In contrast to effective duration, key rate durations help identify "shaping risk" for a bond—that is, a bond's sensitivity to changes in the shape of the benchmark yield curve (e.g., the yield curve becoming steeper or flatter).

The previous illustration of effective duration assumed a parallel shift of 25 bps at all maturities. However, the analyst may want to know how the price of the callable bond is expected to change if short-term benchmark rates (say, for a current two-year Treasury note with modified duration of 1.9) rise by 25 bps but longer-maturity benchmark rates remain unchanged. This scenario would represent a flattening of the yield curve, given that the yield curve is upward sloping. Using key rate durations, the expected price change would be approximately equal to minus the key rate duration for the short-maturity segment (–1.9) times the 0.0025 interest rate shift at that segment, or –0.475% based on the following formula:

$$\frac{\Delta PV}{PV} = -\text{KeyRateDur}^k \times \Delta r^k$$

Of course, for parallel shifts in the benchmark yield curve, key rate durations will indicate the same interest rate sensitivity as effective duration.

6 PROPERTIES OF BOND DURATION

e explain how a bond's maturity, coupon, and yield level affect its interest rate risk

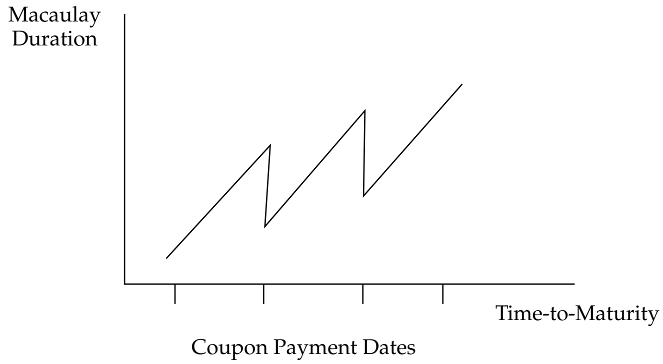
The Macaulay and modified yield duration statistics for a traditional fixed-rate bond are functions of the input variables: the coupon rate or payment per period, the yield-to-maturity per period, the number of periods to maturity (as of the beginning of the period), and the fraction of the period that has gone by. The properties of bond duration are obtained by changing one of these variables while holding the others constant. Because duration is the basic measure of interest rate risk on a fixed-rate bond, these properties are important to understand.

The closed-form formula for Macaulay duration, presented as Equation 4 and again here, is useful in demonstrating the characteristics of the bond duration statistic.

$$\text{MacDur} = \left\{\frac{1+r}{r} - \frac{1+r+\left[N \times (c-r)\right]}{c \times \left[(1+r)^N - 1\right] + r}\right\} - (t/T)$$

The same characteristics hold for modified duration. Consider first the fraction of the period that has gone by (t/T). Macaulay and modified durations depend on the day-count basis used to obtain the yield-to-maturity. The duration of a bond that uses the actual/actual method to count days is slightly different from that of an otherwise comparable bond that uses the 30/360 method. The key point is that for a constant yield-to-maturity (r), the expression in braces is unchanged as time passes during the period. Therefore, the Macaulay duration decreases smoothly as t goes from $t = 0$ to $t = T$, which creates a "saw-tooth" pattern. This pattern for a typical fixed-rate bond is illustrated in Exhibit 6.

Exhibit 6 Macaulay Duration between Coupon Payments with a Constant Yield-to-Maturity



As times passes during the coupon period (moving from right to left in the diagram), the Macaulay duration declines smoothly and then jumps upward after the coupon is paid.

The characteristics of bond duration related to changes in the coupon rate, the yield-to-maturity, and the time-to-maturity are illustrated in Exhibit 7.

Exhibit 7 Properties of the Macaulay Yield Duration

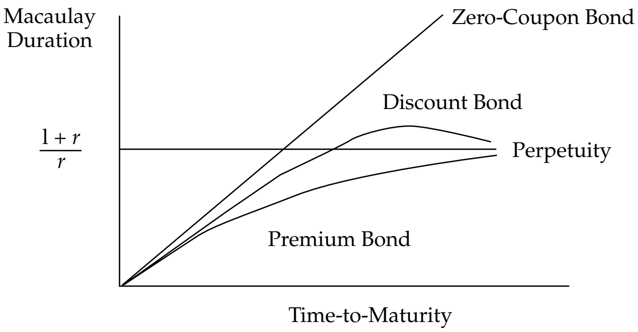


Exhibit 7 shows the graph for coupon payment dates when $t/T = 0$, thus not displaying the saw-tooth pattern between coupon payments. The relationship between the Macaulay duration and the time-to-maturity for a zero-coupon bond is the 45-degree line: MacDur = N when $c = 0$ (and $t/T = 0$). Therefore, the Macaulay duration of a zero-coupon bond is its time-to-maturity.

A **perpetuity** or perpetual bond, which also is called a consol, is a bond that does not mature. There is no principal to redeem. The investor receives a fixed coupon payment forever unless the bond is callable. Non-callable perpetuities are rare, but they have an interesting Macaulay duration: MacDur = $(1 + r)/r$ as N approaches infinity. In effect, the second expression within the braces approaches zero as the number of periods to maturity increases because N in the numerator is a coefficient but N in the denominator is an exponent and the denominator increases faster than the numerator as N grows larger.

Typical fixed-rate coupon bonds with a stated maturity date are portrayed in Exhibit 7 as the premium and discount bonds. The usual pattern is that longer times-to-maturity correspond to higher Macaulay duration statistics. This pattern always holds for bonds trading at par value or at a premium above par. In Equation 4, the second expression within the braces is a positive number for premium and par bonds.

The numerator is positive because the coupon rate (c) is greater than or equal to the yield-to-maturity (r), whereas the denominator is always positive. Therefore, the Macaulay duration is always less than $(1 + r)/r$, and it approaches that threshold from below as the time-to-maturity increases.

The curious result displayed in Exhibit 7 is in the pattern for discount bonds. Generally, the Macaulay duration increases for a longer time-to-maturity. But at some point when the time-to-maturity is high enough, the Macaulay duration exceeds $(1 + r)/r$, reaches a maximum, and then approaches the threshold from above. In Equation 4, such a pattern develops when the number of periods (N) is large and the coupon rate (c) is below the yield-to-maturity (r). Then the numerator of the second expression within the braces can become negative. The implication is that on long-term discount bonds, the interest rate risk can actually be less than on a shorter-term bond, which explains why the word "generally" is needed in describing the maturity effect for the relationship between bond prices and yields-to-maturity. Generally, for the same coupon rate, a longer-term bond has a greater percentage price change than a shorter-term bond when their yields-to-maturity change by the same amount. The exception is when the longer-term bond has a lower duration statistic.

Coupon rates and yields-to-maturity are both inversely related to the Macaulay duration. In Exhibit 7, for the same time-to-maturity and yield-to-maturity, the Macaulay duration is higher for a zero-coupon bond than for a low-coupon bond trading at a discount. Also, the low-coupon bond trading at a discount has a higher duration than a high-coupon bond trading at a premium. Therefore, all else being equal, a lower-coupon bond has a higher duration and more interest rate risk than a higher-coupon bond. The same pattern holds for the yield-to-maturity. A higher yield-to-maturity reduces the weighted average of the time to receipt of cash flow. More weight is on the cash flows received in the near term, and less weight is on the cash flows received in the more-distant future periods if those cash flows are discounted at a higher rate.

In summary, the Macaulay and modified duration statistics for a fixed-rate bond depend primarily on the coupon rate, yield-to-maturity, and time-to-maturity. A higher coupon rate or a higher yield-to-maturity reduces the duration measures. A longer time-to-maturity *usually* leads to a higher duration. It *always* does so for a bond priced at a premium or at par value. But if the bond is priced at a discount, a longer time-to-maturity *might* lead to a lower duration. This situation only occurs if the coupon rate is low (but not zero) relative to the yield and the time-to-maturity is long.

EXAMPLE 10

A hedge fund specializes in investments in emerging market sovereign debt. The fund manager believes that the implied default probabilities are too high, which means that the bonds are viewed as "cheap" and the credit spreads are too high. The hedge fund plans to take a position on one of these available bonds.

Bond	Time-to-Maturity	Coupon Rate	Price	Yield-to-Maturity
(A)	10 years	10%	58.075279	20%
(B)	20 years	10%	51.304203	20%
(C)	30 years	10%	50.210636	20%

The coupon payments are annual. The yields-to-maturity are effective annual rates. The prices are per 100 of par value.

1 Compute the approximate modified duration of each of the three bonds using a 1 bp change in the yield-to-maturity and keeping precision to six decimals (because approximate duration statistics are very sensitive to rounding).

2 Which of the three bonds is expected to have the highest percentage price increase if the yield-to-maturity on each decreases by the same amount—for instance, by 10 bps from 20% to 19.90%?

Solution to 1:

Bond A:

$PV_0 = 58.075279$

$PV_+ = 58.047598$

$$\frac{10}{(1.2001)^1} + \frac{10}{(1.2001)^2} + \cdots + \frac{110}{(1.2001)^{10}} = 58.047598$$

$PV_- = 58.102981$

$$\frac{10}{(1.1999)^1} + \frac{10}{(1.1999)^2} + \cdots + \frac{110}{(1.1999)^{10}} = 58.102981$$

The approximate modified duration of Bond A is 4.768.

$$\text{ApproxModDur} = \frac{58.102981 - 58.047598}{2 \times 0.0001 \times 58.075279} = 4.768$$

Bond B:

$PV_0 = 51.304203$

$PV_+ = 51.277694$

$$\frac{10}{(1.2001)^1} + \frac{10}{(1.2001)^2} + \cdots + \frac{110}{(1.2001)^{20}} = 51.277694$$

$PV_- = 51.330737$

$$\frac{10}{(1.1999)^1} + \frac{10}{(1.1999)^2} + \cdots + \frac{110}{(1.1999)^{20}} = 51.330737$$

The approximate modified duration of Bond B is 5.169.

$$\text{ApproxModDur} = \frac{51.330737 - 51.277694}{2 \times 0.0001 \times 51.304203} = 5.169$$

Bond C:

$PV_0 = 50.210636$

$PV_+ = 50.185228$

$$\frac{10}{(1.2001)^1} + \frac{10}{(1.2001)^2} + \cdots + \frac{110}{(1.2001)^{30}} = 50.185228$$

$PV_- = 50.236070$

$$\frac{10}{(1.1999)^1} + \frac{10}{(1.1999)^2} + \cdots + \frac{110}{(1.1999)^{30}} = 50.236070$$

The approximate modified duration of Bond C is 5.063.

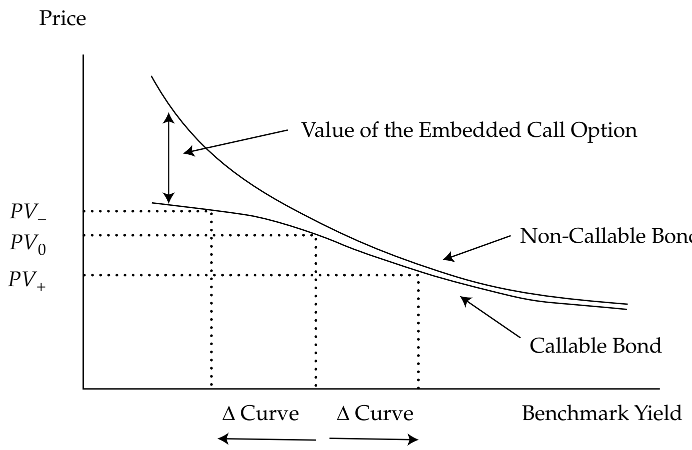
$$\text{ApproxModDur} = \frac{50.236070 - 50.185228}{2 \times 0.0001 \times 50.210636} = 5.063$$

Solution to 2:

Despite the significant differences in times-to-maturity (10, 20, and 30 years), the approximate modified durations on the three bonds are fairly similar (4.768, 5.169, and 5.063). Because the yields-to-maturity are so high, the additional time to receipt of interest and principal payments on the 20- and 30-year bonds have low weight. Nevertheless, Bond B, with 20 years to maturity, has the highest modified duration. If the yield-to-maturity on each is decreased by the same amount—for instance, by 10 bps, from 20% to 19.90%—Bond B would be expected to have the highest percentage price increase because it has the highest modified duration. This example illustrates the relationship between the Macaulay duration and the time-to-maturity on discount bonds in Exhibit 7. The 20-year bond has a higher duration than the 30-year bond.

Callable bonds require the use of effective duration because Macaulay and modified yield duration statistics are not relevant. The yield-to-maturity for callable bonds is not well-defined because future cash flows are uncertain. Exhibit 8 illustrates the impact of the change in the benchmark yield curve (ΔCurve) on the price of a callable bond price compared with that on a comparable non-callable bond. The two bonds have the same credit risk, coupon rate, payment frequency, and time-to-maturity. The vertical axis is the bond price. The horizontal axis is a benchmark yield—for instance, a point on the par curve for government bonds.

Exhibit 8 Interest Rate Risk Characteristics of a Callable Bond

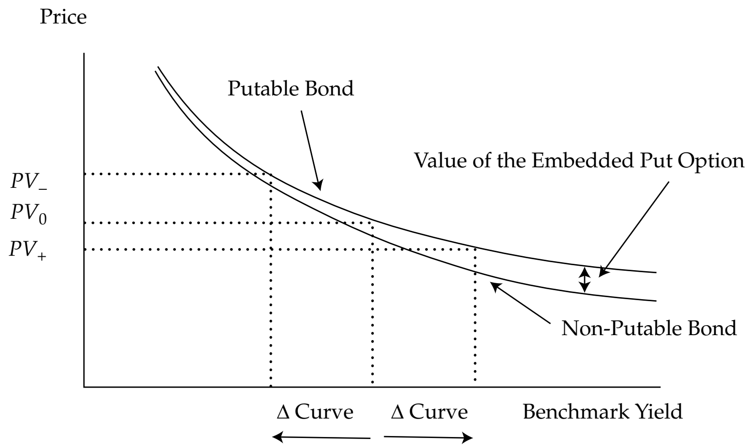


As shown in Exhibit 8, the price of the non-callable bond is always greater than that of the callable bond with otherwise identical features. The difference is the value of the embedded call option. Recall that the call option is an option to the issuer and not the holder of the bond. When interest rates are high compared with the coupon rate, the value of the call option is low. When rates are low, the value of the call option is much greater because the issuer is more likely to exercise the option to refinance the debt at a lower cost of funds. The investor bears the "call risk" because if the bond is called, the investor must reinvest the proceeds at a lower interest rate.

Exhibit 8 shows the inputs for calculating the effective duration of the callable bond. The entire benchmark curve is raised and lowered by the same amount, ΔCurve. The key point is that when benchmark yields are high, the effective durations of the callable and non-callable bonds are very similar. Although the exhibit does not illustrate it, the slopes of the lines tangent to the price–yield curve are about the same in such a situation. But when interest rates are low, the effective duration of the callable bond is lower than that of the otherwise comparable non-callable bond. That is because the callable bond price does not increase as much when benchmark yields fall. The slope of the line tangent to the price–yield curve would be flatter. The presence of the call option limits price appreciation. Therefore, an embedded call option reduces the effective duration of the bond, especially when interest rates are falling and the bond is more likely to be called. The lower effective duration can also be interpreted as a shorter expected life—the weighted average of time to receipt of cash flow is reduced.

Exhibit 9 considers another embedded option—a put option.

Exhibit 9 Interest Rate Risk Characteristics of a Putable Bond



A putable bond allows the investor to sell the bond back to the issuer prior to maturity, usually at par value, which protects the investor from higher benchmark yields or credit spreads that otherwise would drive the bond to a discounted price. Therefore, the price of a putable bond is always higher than that of an otherwise comparable non-putable bond. The price difference is the value of the embedded put option.

An embedded put option reduces the effective duration of the bond, especially when rates are rising. If interest rates are low compared with the coupon rate, the value of the put option is low and the impact of a change in the benchmark yield on the bond's price is very similar to the impact on the price of a non-putable bond. But when benchmark interest rates rise, the put option becomes more valuable to the investor. The ability to sell the bond at par value limits the price depreciation as rates rise. In summary, the presence of an embedded option reduces the sensitivity of the bond price to changes in the benchmark yield curve, assuming no change in credit risk.

7 DURATION OF A BOND PORTFOLIO

f calculate the duration of a portfolio and explain the limitations of portfolio duration

Similar to equities, bonds are typically held in a portfolio. There are two ways to calculate the duration of a bond portfolio: (1) the weighted average of time to receipt of the *aggregate* cash flows, and (2) the weighted average of the individual bond durations that comprise the portfolio. The first method is the theoretically correct approach, but it is difficult to use in practice. The second method is commonly used by fixed-income portfolio managers, but it has its own limitations. The differences in these two methods to compute portfolio duration can be examined with a numerical example.

Suppose an investor holds the following portfolio of two *zero-coupon* bonds:

Bond	Maturity	Price	Yield	Macaulay Duration	Modified Duration	Par Value	Market Value	Weight
(X)	1 year	98.00	2.0408%	1	0.980	10,000,000	9,800,000	0.50
(Y)	30 years	9.80	8.0503%	30	27.765	100,000,000	9,800,000	0.50

The prices are per 100 of par value. The yields-to-maturity are effective annual rates. The total market value for the portfolio is 19,600,000. The portfolio is evenly weighted in terms of market value between the two bonds.

The first approach views the portfolio as a series of aggregated cash flows. Its **cash flow yield** is 7.8611%. A cash flow yield is the internal rate of return on a series of cash flows, usually used on a complex security such as a mortgage-backed bond (using projected cash flows based on a model of prepayments as a result of refinancing) or a portfolio of fixed-rate bonds. It is the solution for r in the following equation.

$$19{,}600{,}000 = \frac{10{,}000{,}000}{(1+r)^1} + \frac{0}{(1+r)^2} + \cdots + \frac{0}{(1+r)^{29}} + \frac{100{,}000{,}000}{(1+r)^{30}}, \quad r = 0.078611$$

The Macaulay duration of the portfolio in this approach is the weighted average of the times to receipt of aggregated cash flow. The cash flow yield is used to obtain the weights. This calculation is similar to Equation 1, and the portfolio duration is 16.2825.

$$\text{MacDur} = \left[\frac{\dfrac{1 \times 10{,}000{,}000}{(1.078611)^1} + \dfrac{30 \times 100{,}000{,}000}{(1.078611)^{30}}}{\dfrac{10{,}000{,}000}{(1.078611)^1} + \dfrac{100{,}000{,}000}{(1.078611)^{30}}}\right] = 16.2825$$

There are just two future cash flows in the portfolio—the redemption of principal on the two zero-coupon bonds. In more complex portfolios, a series of coupon and principal payments may occur on some dates, with an aggregated cash flow composed of coupon interest on some bonds and principal on those that mature.

The modified duration of the portfolio is the Macaulay duration divided by one plus the cash flow yield per period (here, the periodicity is 1).

$$\text{ModDur} = \frac{16.2825}{1.078611} = 15.0958$$

The modified duration for the portfolio is 15.0958. That statistic indicates the percentage change in the market value given a change in the cash flow yield. If the cash flow yield increases or decreases by 100 bps, the market value of the portfolio is expected to decrease or increase by about 15.0958%.

Although this approach is theoretically correct, it is difficult to use in practice. First, the cash flow yield is not commonly calculated for bond portfolios. Second, the amount and timing of future coupon and principal payments are uncertain if the portfolio contains callable or putable bonds or floating-rate notes. Third, interest rate risk is usually expressed as a change in benchmark interest rates, not as a change in the cash flow yield. Fourth, the change in the cash flow yield is not necessarily the same amount as the change in the yields-to-maturity on the individual bonds. For instance, if the yields-to-maturity on the two zero-coupon bonds in this portfolio both increase or decrease by 10 bps, the cash flow yield increases or decreases by only 9.52 bps.

In practice, the second approach to portfolio duration is commonly used. The Macaulay and modified durations for the portfolio are calculated as the weighted average of the statistics for the individual bonds. The shares of overall portfolio market value are the weights. This weighted average approximates the theoretically correct portfolio duration, which is obtained using the first approach. This approximation becomes more accurate when the differences in the yields-to-maturity on the bonds in the portfolio are smaller. When the yield curve is flat, the two approaches produce the same portfolio duration.

Given the equal "50/50" weights in this simple numerical example, this version of portfolio duration is easily computed.

$$\text{Average Macaulay duration} = (1 \times 0.50) + (30 \times 0.50) = 15.50$$

$$\text{Average modified duration} = (0.980 \times 0.50) + (27.765 \times 0.50) = 14.3725$$

Note that 0.980 = 1/1.020404 and 27.765 = 30/1.080503. An advantage of the second approach is that callable bonds, putable bonds, and floating-rate notes can be included in the weighted average using the effective durations for these securities.

The main advantage to the second approach is that it is easily used as a measure of interest rate risk. For instance, if the yields-to-maturity on the bonds in the portfolio increase by 100 bps, the estimated drop in the portfolio value is 14.3725%. However, this advantage also indicates a limitation: This measure of portfolio duration implicitly assumes a **parallel shift** in the yield curve. A parallel yield curve shift implies that all rates change by the same amount in the same direction. In reality, interest rate changes frequently result in a steeper or flatter yield curve. Yield volatility is discussed later in this reading.

EXAMPLE 11

An investment fund owns the following portfolio of three fixed-rate government bonds:

	Bond A	Bond B	Bond C
Par value	EUR25,000,000	EUR25,000,000	EUR50,000,000
Coupon rate	9%	11%	8%
Time-to-maturity	6 years	8 years	12 years
Yield-to-maturity	9.10%	9.38%	9.62%
Market value	EUR24,886,343	EUR27,243,887	EUR44,306,787
Macaulay duration	4.761	5.633	7.652

The total market value of the portfolio is EUR96,437,017. Each bond is on a coupon date so that there is no accrued interest. The market values are the full prices given the par value. Coupons are paid semiannually. The yields-to-maturity are stated on a semiannual bond basis, meaning an annual rate for a periodicity of 2. The Macaulay durations are annualized.

1 Calculate the average (annual) modified duration for the portfolio using the shares of market value as the weights.

2 Estimate the percentage loss in the portfolio's market value if the (annual) yield-to-maturity on each bond goes up by 20 bps.

Solution to 1:

The average (annual) modified duration for the portfolio is 6.0495.

$$\left(\frac{4.761}{1+\frac{0.0910}{2}} \times \frac{24{,}886{,}343}{96{,}437{,}017}\right) + \left(\frac{5.633}{1+\frac{0.0938}{2}} \times \frac{27{,}243{,}887}{96{,}437{,}017}\right) +$$

$$\left(\frac{7.652}{1+\frac{0.0962}{2}} \times \frac{44{,}306{,}787}{96{,}437{,}017}\right) = 6.0495$$

Note that the annual modified duration for each bond is the annual Macaulay duration, which is given, divided by one plus the yield-to-maturity per semi-annual period.

Solution to 2:

The estimated decline in market value if each yield rises by 20 bps is 1.21%: $-6.0495 \times 0.0020 = -0.0121$.

8 MONEY DURATION AND THE PRICE VALUE OF A BASIS POINT

g calculate and interpret the money duration of a bond and price value of a basis point (PVBP)

Modified duration is a measure of the *percentage price change* of a bond given a change in its yield-to-maturity. A related statistic is **money duration**. The money duration of a bond is a measure of the *price change* in units of the currency in which the bond is denominated. The money duration can be stated per 100 of par value or in terms of the actual position size of the bond in the portfolio. In the United States, money duration is commonly called "dollar duration."

Money duration (MoneyDur) is calculated as the annual modified duration times the full price (PV^{Full}) of the bond, including accrued interest.

$$\text{MoneyDur} = \text{AnnModDur} \times PV^{Full} \tag{11}$$

The estimated change in the bond price in currency units is calculated using Equation 12, which is very similar to Equation 6. The difference is that for a given change in the annual yield-to-maturity (ΔYield), modified duration estimates the percentage price change and money duration estimates the change in currency units.

$$\Delta PV^{Full} \approx -\text{MoneyDur} \times \Delta\text{Yield} \tag{12}$$

For a theoretical example of money duration, consider the 6% semiannual coupon payment bond that matures on 14 February 2027 and is priced to yield 6.00% for settlement on 11 April 2019. The full price of the bond is 100.940423 per 100 of par value, and the annual modified duration is 6.1268. Suppose that a Nairobi based life insurance company has a position in the bond for a par value of KES100,000,000. The market value of the investment is KES 100,940,423. The money duration of this bond is KES 618,441,784 (= 6.1268 × KES 100,940,423). Therefore, if the yield-to-maturity rises by 100 bps—from 6.00% to 7.00%—the expected loss is approximately –KES 6,184,418 (= –KES 618,441,784 × 0.0100). On a percentage basis, that expected loss is approximately 6.1268%. The "convexity adjustment" introduced in the next section makes these estimates more accurate.

Another version of money duration is the value of one basis point in price terms. The **price value of a basis point** (or PVBP) is an estimate of the change in the full bond price given a 1 bp change in the yield-to-maturity. The PVBP can be calculated using a formula similar to that for the approximate modified duration. Equation 13 is the formula for the PVBP.

$$\text{PVBP} = \frac{(PV_-) - (PV_+)}{2} \tag{13}$$

PV_- and PV_+ are the full prices calculated by decreasing and increasing the yield-to-maturity by 1 bp. The PVBP is also called the "PV01," standing for the "price value of an 01" or "present value of an 01," where "01" means 1 bp. In the United States, it is commonly called the "DV01," or the "dollar value of a 01." The PVBP is particularly useful for bonds where future cash flows are uncertain, such as callable bonds. A related statistic called a "basis point value" (BPV) is simply the money duration times 0.0001 (1 bp).

For a numerical example of the PVBP calculation, consider the 2.875% semiannual coupon payment US Treasury note that matures on 15 May 2028. In Exhibit 5, the PVBP for the Treasury note is shown to be 0.08540. Its yield-to-maturity is 2.849091%, and the settlement date is 59 days into a 184-day period. To confirm this result, calculate the new prices by increasing and decreasing the yield-to-maturity. First, increase the yield by 1 bp (0.01%), from 2.849091% to 2.859091%, to solve for a PV_+ of 100.594327.

$$PV_+ = \left[\frac{1.4375}{\left(1+\frac{0.02859091}{2}\right)^1} + \cdots + \frac{101.4375}{\left(1+\frac{0.02859091}{2}\right)^{20}}\right] \times \left(1+\frac{0.02859091}{2}\right)^{59/184}$$
$$= 101.594327$$

Then, decrease the yield-to-maturity by 1 bp, from 2.849091% to 2.839091%, to solve for a PV_- of 100.765123.

$$PV_- = \left[\frac{1.4375}{\left(1+\frac{0.02839091}{2}\right)^1} + \cdots + \frac{101.4375}{\left(1+\frac{0.02839091}{2}\right)^{20}}\right] \times \left(1+\frac{0.02839091}{2}\right)^{59/184}$$
$$= 100.765123$$

The PVBP is obtained by substituting these results into Equation 13.

$$\text{PVBP} = \frac{100.765123 - 100.594327}{2} = 0.08540$$

Another money duration statistic reported on the Bloomberg YAS page is "risk." It is shown to be 8.540. Bloomberg's risk statistic is simply the PVBP (or PV01) times 100.

EXAMPLE 12

A life insurance company holds a USD10 million (par value) position in a 5.95% Dominican Republic bond that matures on 25 January 2027. The bond is priced (flat) at 101.996 per 100 of par value to yield 5.6511% on a street-convention semiannual bond basis for settlement on 24 July 2018. The total market value of the position, including accrued interest, is USD10,495,447, or 101.95447 per 100 of par value. The bond's (annual) Macaulay duration is 6.622.

1. Calculate the money duration per 100 in par value for the sovereign bond.
2. Using the money duration, estimate the loss on the position for each 1 bp increase in the yield-to-maturity for that settlement date.

Solution to 1:

The money duration is the annual modified duration times the full price of the bond per 100 of par value.

$$\left(\frac{6.622}{1+\frac{0.056511}{2}}\right) \times \text{USD}104.95447 = \text{USD}675.91$$

Solution to 2:

For each 1 bp increase in the yield-to-maturity, the loss is estimated to be USD 0.067591 per 100 of par value: USD 675.91 × 0.0001 = USD 0.067591.

Given a position size of USD 10 million in par value, the estimated loss per basis-point increase in the yield is USD 6,759.10. The money duration is per 100 of par value, so the position size of USD10 million is divided by USD 100.

$$\text{USD}0.067591 \times \frac{\text{USD}10{,}000{,}000}{\text{USD}100} = \text{USD}6{,}759.10$$

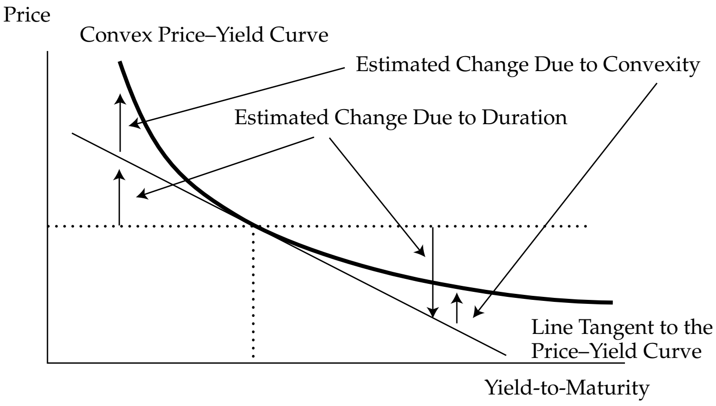
9 BOND CONVEXITY

h **calculate and interpret approximate convexity and compare approximate and effective convexity**

i **calculate the percentage price change of a bond for a specified change in yield, given the bond's approximate duration and convexity**

Modified duration measures the primary effect on a bond's percentage price change given a change in the yield-to-maturity. A secondary effect is measured by the convexity statistic, which is illustrated in Exhibit 10 for a traditional (option-free) fixed-rate bond.

Exhibit 10 Convexity of a Traditional (Option-Free) Fixed-Rate Bond



The true relationship between the bond price and the yield-to-maturity is the curved (convex) line shown in Exhibit 10. This curved line shows the actual bond price given its market discount rate. Duration (in particular, money duration) estimates the change in the bond price along the straight line that is tangent to the curved line. For small yield-to-maturity changes, there is little difference between the lines. But for larger changes, the difference becomes significant.

The convexity statistic for the bond is used to improve the estimate of the percentage price change provided by modified duration alone. Equation 14 is the convexity-adjusted estimate of the percentage change in the bond's full price.

$$\%\Delta PV^{Full} \approx$$
$$\left(-\text{AnnModDur} \times \Delta\text{Yield}\right) + \left[\frac{1}{2} \times \text{AnnConvexity} \times \left(\Delta\text{Yield}\right)^2\right] \quad (14)$$

The first bracketed expression, the "first-order" effect, is the same as Equation 6. The (annual) modified duration, AnnModDur, is multiplied by the change in the (annual) yield-to-maturity, ΔYield. The second bracketed expression, the "second-order" effect, is the **convexity adjustment**. The convexity adjustment is the annual convexity statistic, AnnConvexity, times one-half, multiplied by the change in the yield-to-maturity *squared*. This additional term is a positive amount on a traditional (option-free) fixed-rate bond for either an increase or decrease in the yield. In Exhibit 10, this amount adds to the linear estimate provided by the duration alone, which brings the adjusted estimate very close to the actual price on the curved line. But it still is an estimate, so the ≈ sign is used.

Similar to the Macaulay and modified durations, the annual convexity statistic can be calculated in several ways. It can be calculated using tables, such as Exhibits 2 and 3. It also is possible to derive a closed-form equation for the convexity of a fixed-rate bond on and between coupon payment dates using calculus and algebra (see D. Smith, 2014). But like modified duration, convexity can be approximated with accuracy. Equation 15 is the formula for the approximate convexity statistic, ApproxCon.

$$\text{ApproxCon} = \frac{\left(PV_-\right) + \left(PV_+\right) - \left[2 \times \left(PV_0\right)\right]}{\left(\Delta\text{Yield}\right)^2 \times \left(PV_0\right)} \quad (15)$$

This equation uses the same inputs as Equation 7 for ApproxModDur. The new price when the yield-to-maturity is increased is PV_+. The new price when the yield is decreased by the same amount is PV_-. The original price is PV_0. These are the full prices, including accrued interest, for the bond.

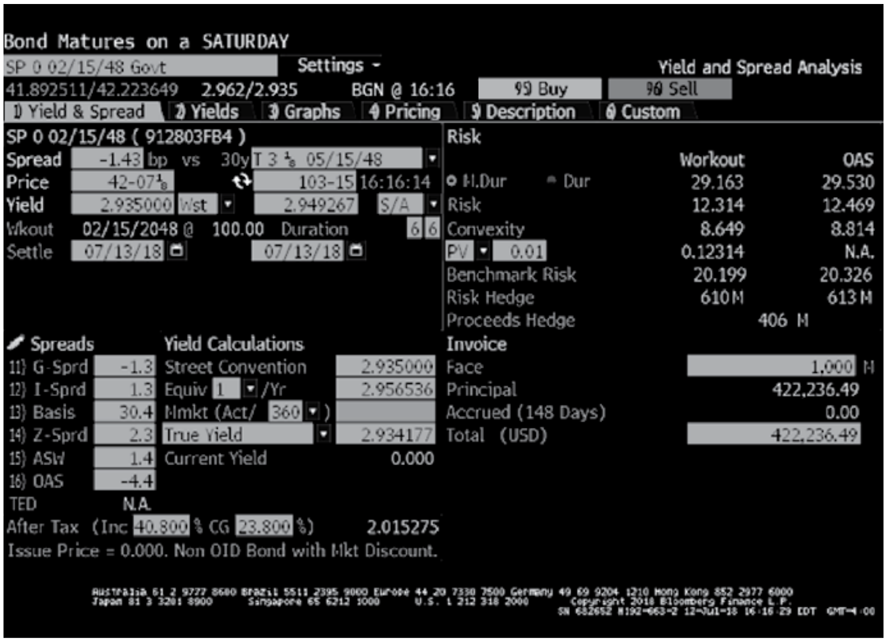
The accuracy of this approximation can be demonstrated with the special case of a zero-coupon bond. The absence of coupon payments simplifies the interest rate risk measures. The Macaulay duration of a zero-coupon bond is $N - t/T$ in terms of periods to maturity. The exact convexity statistic of a zero-coupon bond, also in terms of periods, is calculated with Equation 16.

$$\text{Convexity (of a zero-coupon bond)} = \frac{\left[N - (t/T)\right] \times \left[N + 1 - (t/T)\right]}{(1+r)^2} \tag{16}$$

N is the number of periods to maturity as of the beginning of the current period, t/T is the fraction of the period that has gone by, and r is the yield-to-maturity per period.

For an example of this calculation, consider a long-term, zero-coupon US Treasury bond. The bond's Bloomberg YAS page is shown in Exhibit 11.

Exhibit 11 Bloomberg YAS Page for the Zero-Coupon US Treasury Bond



The bond matures on 15 February 2048 and its asked price was 42.223649 per 100 of par value for settlement on 13 July 2018. Its yield-to-maturity was 2.935% stated on a street-convention semiannual bond basis. Even though it is a zero-coupon bond, its yield-to-maturity is based on the actual/actual day-count convention. That settlement date was 148 days into a 181-day period. The annual modified duration was 29.163.

For this bond, $N = 60$, $t/T = 148/181$, and $r = 0.02935/2$. Entering these variables into Equation 16 produces a convexity of 3,459.45 in terms of semiannual periods.

$$\frac{\left[60 - (148/181)\right] \times \left[60 + 1 - (148/181)\right]}{\left(1 + \dfrac{0.02935}{2}\right)^2} = 3{,}459.45$$

As with the other statistics, convexity is annualized in practice and for use in the convexity adjustment in Equation 14. It is divided by the periodicity *squared*. The yield-to-maturity on this zero-coupon bond is stated on a semiannual bond basis, meaning a periodicity of 2. Therefore, the annualized convexity statistic is 864.9.

$$\frac{3{,}459.45}{4} = 864.9$$

For example, suppose that the yield-to-maturity is expected to fall by 10 bps, from 2.935% to 2.835%. Given the (annual) modified duration of 29.163 and (annual) convexity of 864.9, the expected percentage price gain is 2.9595%.

$$\begin{aligned}\%\Delta PV^{Full} &\approx \left[-29.163 \times -0.0010\right] + \left[\frac{1}{2} \times 864.9 \times (-0.0010)^2\right] \\ &= 0.029163 + 0.000432 \\ &= 0.029595\end{aligned}$$

Modified duration alone (under)estimates the gain to be 2.9163%. The convexity adjustment adds 4.32 bps.

The long-term, zero-coupon bond of Exhibit 11 demonstrates the difference between *yield* duration and convexity and *curve* duration and convexity, even on an option-free bond. Its modified duration is 29.163, whereas its effective duration is 29.530. Its yield convexity is reported on the Bloomberg page to be 8.649, and its effective convexity is 8.814. (Note that although Bloomberg scales the convexity statistics by dividing by 100, either raw or scaled convexity figures are acceptable in practice.) In general, the differences are heightened when the benchmark yield curve is not flat, when the bond has a long time-to-maturity, and the bond is priced at a significant discount or premium.

To obtain the ApproxCon for this long-term, zero-coupon bond, calculate PV_0, PV_+, and PV_- for yields-to-maturity of 2.935%, 2.945%, and 2.925%, respectively. For this exercise, ΔYield = 0.0001.

$$PV_0 = \frac{100}{\left(1 + \frac{0.02935}{2}\right)^{60}} \times \left(1 + \frac{0.02935}{2}\right)^{148/181} = 42.223649$$

$$PV_+ = \frac{100}{\left(1 + \frac{0.02945}{2}\right)^{60}} \times \left(1 + \frac{0.02945}{2}\right)^{148/181} = 42.100694$$

$$PV_- = \frac{100}{\left(1 + \frac{0.02925}{2}\right)^{60}} \times \left(1 + \frac{0.02925}{2}\right)^{148/181} = 42.346969$$

Using these results, first calculate ApproxModDur using Equation 7 to confirm that these inputs are correct. In Exhibit 11, modified duration is stated to be 29.163.

$$\text{ApproxModDur} = \frac{42.346969 - 42.100694}{2 \times 0.0001 \times 42.223649} = 29.163$$

Using Equation 15, ApproxCon is 864.9.

$$\text{ApproxCon} = \frac{42.346969 + 42.100694 - (2 \times 42.223649)}{(0.0001)^2 \times 42.223649} = 864.9$$

This result, 864.9, is an approximation for *annualized* convexity. The number of periods in the year is included in the price calculations. This approximation in this example is the same as the "exact" result using the closed-form equation for the special case of the zero-coupon bond. Any small difference is not likely to be meaningful for practical applications.

Because this is an individual zero-coupon bond, it is easy to calculate the new price if the yield-to-maturity does go down by 50 bps, to 2.435%.

$$\frac{100}{\left(1+\frac{0.02435}{2}\right)^{60}} \times \left(1+\frac{0.02435}{2}\right)^{148/181} = 48.860850$$

Therefore, the actual percentage price increase is 15.7192%.

$$\frac{48.860850 - 42.223649}{42.223649} = 0.157192$$

The convexity-adjusted estimate is 15.6626%.

$$\begin{aligned}\%\Delta PV^{Full} &\approx (-29.163 \times -0.0050) + \left[\frac{1}{2} \times 864.9 \times (-0.0050)^2\right] \\ &= 0.145815 + 0.010811 \\ &= 0.156626\end{aligned}$$

EXAMPLE 13

A Dutch bank holds a large position in a zero-coupon Federal Republic of Germany government bond maturing on 11 April 2025. The yield-to-maturity is –0.72% for settlement on 11 May 2020, stated as an effective annual rate on an Actual/Actual basis. That settlement date is 30 days into the 365-day year using this day count method.

1. Calculate the full price of the bond per 100 of par value.
2. Calculate the approximate modified duration and approximate convexity using a 1 bp increase and decrease in the yield-to-maturity.
3. Calculate the estimated convexity-adjusted percentage price change resulting from a 100 bp increase in the yield-to-maturity.
4. Compare the estimated percentage price change with the actual change, assuming the yield-to-maturity jumps 100 bps to 0.28% on that settlement date.

Solution to 1:

There are five years from the beginning of the current period on 11 April 2020 to maturity on 11 April 2025.

The full price of the bond is 103.617526 per 100 of par value. Note that

$$1 + r = 1 + (-0.0072) = 0.9928$$

$$PV_0 = \left[\frac{100}{(0.9928)^5}\right] \times (0.9928)^{30/365} = 103.617526$$

Solution to 2:

PV_+ = 103.566215, and PV_- = 103.668868.

$$PV_+ = \left[\frac{100}{(0.9929)^5}\right] \times (0.9929)^{30/365} = 103.566215$$

$$PV_- = \left[\frac{100}{(0.9927)^5}\right] \times (0.9927)^{30/365} = 103.668868$$

The approximate modified duration is 4.9535.

$$\text{ApproxModDur} = \frac{103.668868 - 103.566215}{2 \times 0.0001 \times 103.617526} = 4.9535$$

The approximate convexity is 29.918.

$$\text{ApproxCon} = \frac{103.668868 + 103.566215 - (2 \times 103.617526)}{(0.0001)^2 \times 103.617526} = 29.918$$

Solution to 3:

The convexity-adjusted percentage price drop resulting from a 100 bp increase in the yield-to-maturity is estimated to be 4.80391%. Modified duration alone estimates the percentage drop to be 4.9535%. The convexity adjustment adds 14.96 bps.

$$\begin{aligned}\%\Delta PV^{Full} &\approx -(4.9535 \times 0.0100) + \left[\frac{1}{2} \times 29.918 \times (-0.0100)^2\right] \\ &= -0.049535 + 0.001496 \\ &= -0.0480391\end{aligned}$$

Solution to 4:

The new full price if the yield-to-maturity goes from –0.72% to 0.28% on that settlement date is 98.634349.

$$PF^{Full} = \left[\frac{100}{(1.0028)^5}\right] \times (1.0028)^{30/365} = 98.634349$$

$$\%\Delta PV^{Full} = \frac{98.634349 - 103.617526}{103.617526} = -0.04809203$$

The actual percentage change in the bond price is –4.809203%. The convexity-adjusted estimate is –4.80391%, whereas the estimated change using modified duration alone is –4.9535%.

The money duration of a bond indicates the first-order effect on the full price of a bond in units of currency given a change in the yield-to-maturity. The **money convexity** statistic (MoneyCon) is the second-order effect. The money convexity of the bond is the annual convexity multiplied by the full price, such that

$$\Delta PV^{Full} \approx -(\text{MoneyDur} \times \Delta\text{Yield}) + \left[\frac{1}{2} \times \text{MoneyCon} \times (\Delta\text{Yield})^2\right] \quad (17)$$

For a money convexity example, consider again the Nairobi-based life insurance company that has a KES100 million position in the 6.00% bond that matures on 14 February 2027. Previously, using the money duration alone, the estimated loss was KES6,184,418 if the yield-to-maturity increased by 100 bps. The money duration for the position is KES618,441,784. That estimation is improved by including the convexity adjustment. Given the approximate modified duration of 6.1268 for a 5 bp change in the yield-to-maturity (ΔYield = 0.0005) and given that PV_0 = 100.940423, PV_+ = 100.631781, and PV_- = 101.250227, we use Equation 15 to calculate the approximate convexity:

$$\text{ApproxCon} = \frac{101.250227 + 100.631781 - (2 \times 100.940423)}{(0.0005)^2 \times 100.940423} = 46.047$$

The money convexity is 46.047 times the market value of the position, KES100,940,423. The convexity-adjusted loss given a 100 bp jump in the yield-to-maturity is KES5,952,018:

$$-\left[(6.1268 \times \text{KES}100{,}940{,}423) \times 0.0100\right] +$$

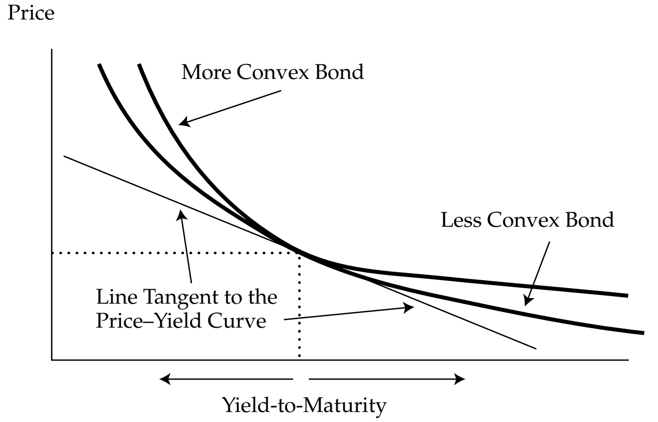
$$\left[\frac{1}{2} \times (46.047 \times \text{KES}100{,}940{,}423) \times (0.0100)^2\right]$$

$$= -\text{KES}6{,}184{,}418 + \text{KES}232{,}400$$

$$= -\text{KES}5{,}952{,}018$$

The factors that lead to greater convexity are the same as for duration. A fixed-rate bond with a longer time-to-maturity, a lower coupon rate, and a lower yield-to-maturity has greater convexity than a bond with a shorter time-to-maturity, a higher coupon rate, and a higher yield-to-maturity. Another factor is the dispersion of cash flows, meaning the degree to which payments are spread out over time. If two bonds have the same duration, the one that has the greater dispersion of cash flows has the greater convexity. The positive attributes of greater convexity for an investor are shown in Exhibit 12.

Exhibit 12 The Positive Attributes of Greater Bond Convexity on a Traditional (Option-Free) Bond



The two bonds in Exhibit 12 are assumed to have the same price, yield-to-maturity, and modified duration. Therefore, they share the same line tangent to their price–yield curves. The benefit of greater convexity occurs when their yields-to-maturity change. For the same decrease in yield-to-maturity, the more convex bond *appreciates more* in price. And for the same increase in yield-to-maturity, the more convex bond *depreciates less* in price. The conclusion is that the more convex bond outperforms the less convex bond in both bull (rising price) and bear (falling price) markets. This conclusion assumes, however, that this positive attribute is not "priced into" the bond. To the extent that it is included, the more convex bond would have a higher price (and lower yield-to-maturity). That does not diminish the value of convexity. It only suggests that the investor has to pay for it. As economists say, "There is no such thing as a free lunch."

EXAMPLE 14

The investment manager for a UK defined-benefit pension scheme is considering two bonds about to be issued by a large life insurance company. The first is a 30-year, 4% semiannual coupon payment bond. The second is a 100-year, 4% semiannual coupon payment "century" bond. Both bonds are expected to trade at par value at issuance.

Calculate the approximate modified duration and approximate convexity for each bond using a 5 bp increase and decrease in the annual yield-to-maturity. Retain accuracy to six decimals per 100 of par value.

Solution:

In the calculations, the yield per semiannual period goes up by 2.5 bps to 2.025% and down by 2.5 bps to 1.975%. The 30-year bond has an approximate modified duration of 17.381 and an approximate convexity of 420.80.

$$PV_+ = \frac{2}{(1.02025)^1} + \cdots + \frac{102}{(1.02025)^{60}} = 99.136214$$

$$PV_- = \frac{2}{(1.01975)^1} + \cdots + \frac{102}{(1.01975)^{60}} = 100.874306$$

$$\text{ApproxModDur} = \frac{100.874306 - 99.136214}{2 \times 0.0005 \times 100} = 17.381$$

$$\text{ApproxCon} = \frac{100.874306 + 99.136214 - (2 \times 100)}{(0.0005)^2 \times 100} = 420.80$$

The 100-year century bond has an approximate modified duration of 24.527 and an approximate convexity of 1,132.88.

$$PV_+ = \frac{2}{(1.02025)^1} + \cdots + \frac{102}{(1.02025)^{200}} = 98.787829$$

$$PV_- = \frac{2}{(1.01975)^1} + \cdots + \frac{102}{(1.01975)^{200}} = 101.240493$$

$$\text{ApproxModDur} = \frac{101.240493 - 98.787829}{2 \times 0.0005 \times 100} = 24.527$$

$$\text{ApproxCon} = \frac{101.240493 + 98.787829 - (2 \times 100)}{(0.0005)^2 \times 100} = 1{,}132.88$$

The century bond offers a higher modified duration, 24.527 compared with 17.381, and a much greater degree of convexity, 1,132.88 compared with 420.80.

In the same manner that the primary, or first-order, effect of a shift in the benchmark yield curve is measured by effective duration, the secondary, or second-order, effect is measured by **effective convexity**. The effective convexity of a bond is a *curve convexity* statistic that measures the secondary effect of a change in a benchmark yield curve. A pricing model is used to determine the new prices when the benchmark curve is shifted upward (PV_+) and downward (PV_-) by the same amount (ΔCurve). These changes are made holding other factors constant—for example, the credit spread. Then, Equation 18 is used to calculate the effective convexity (EffCon) given the initial price (PV_0).

$$\text{EffCon} = \frac{\left[(PV_-) + (PV_+)\right] - \left[2 \times (PV_0)\right]}{(\Delta\text{Curve})^2 \times (PV_0)} \tag{18}$$

This equation is very similar to Equation 15, for approximate *yield* convexity. The difference is that in Equation 15, the denominator includes the change in the yield-to-maturity squared, $(\Delta\text{Yield})^2$. Here, the denominator includes the change in the benchmark yield curve squared, $(\Delta\text{Curve})^2$.

Consider again the callable bond example in our initial discussion of effective duration. It is assumed that an option-pricing model is used to generate these callable bond prices: PV_0 = 101.060489, PV_+ = 99.050120, PV_- = 102.890738, and ΔCurve = 0.0025. The effective duration for the callable bond is 7.6006.

$$\text{EffDur} = \frac{102.890738 - 99.050120}{2 \times 0.0025 \times 101.060489} = 7.6006$$

Using these inputs in Equation 18, the effective convexity is –285.17.

$$\text{EffCon} = \frac{102.890738 + 99.050120 - (2 \times 101.060489)}{(0.0025)^2 \times 101.060489} = -285.17$$

Negative convexity, which could be called "concavity," is an important feature of callable bonds. Putable bonds, on the other hand, always have positive convexity. As a second-order effect, effective convexity indicates the change in the first-order effect (i.e., effective duration) as the benchmark yield curve is changed. In Exhibit 8, as the benchmark yield goes down, the slope of the line tangent to the curve for the non-callable bond steepens, which indicates positive convexity. But the slope of the line tangent to the callable bond flattens as the benchmark yield goes down. Technically, it reaches an inflection point, which is when the effective convexity shifts from positive to negative.

In summary, when the benchmark yield is high and the value of the embedded call option is low, the callable and the non-callable bonds experience very similar effects from interest rate changes. They both have positive convexity. But as the benchmark yield is reduced, the curves diverge. At some point, the callable bond moves into the range of negative convexity, which indicates that the embedded call option has more value to the issuer and is more likely to be exercised. This situation limits the potential price appreciation of the bond arising from lower interest rates, whether because of a lower benchmark yield or a lower credit spread.

Another way to understand why a callable bond can have negative convexity is to rearrange Equation 18.

$$\text{EffCon} = \frac{\left[(PV_-) - (PV_0)\right] - \left[(PV_0) - (PV_+)\right]}{(\Delta\text{Curve})^2 \times (PV_0)}$$

In the numerator, the first bracketed expression is the increase in price when the benchmark yield curve is lowered. The second expression is the decrease in price when the benchmark yield curve is raised. On a non-callable bond, the increase is always larger than the decrease (in absolute value). This result is the "convexity effect" for the relationship between bond prices and yields-to-maturity. On a callable bond, the increase can be smaller than the decrease (in absolute value). That creates negative convexity, as illustrated in Exhibit 8.

INVESTMENT HORIZON, MACAULAY DURATION AND INTEREST RATE RISK 10

j describe how the term structure of yield volatility affects the interest rate risk of a bond

k describe the relationships among a bond's holding period return, its duration, and the investment horizon

This section explores the effect of yield volatility on the investment horizon, and on the interaction between the investment horizon, market price risk, and coupon reinvestment risk.

10.1 Yield Volatility

An important aspect in understanding the interest rate risk and return characteristics of an investment in a fixed-rate bond is the time horizon. This section considers a short-term horizon. A primary concern for the investor is the change in the price of the bond given a sudden (i.e., same-day) change in its yield-to-maturity. The accrued interest does not change, so the impact of the change in the yield is on the flat price of the bond. Next, we consider a long-term horizon. The reinvestment of coupon interest then becomes a key factor in the investor's horizon yield.

Bond duration is the primary measure of risk arising from a change in the yield-to-maturity. Convexity is the secondary risk measure. In the discussion of the impact on the bond price, the phrase "for a *given* change in the yield-to-maturity" is used repeatedly. For instance, the given change in the yield-to-maturity could be 1 bp, 25 bps, or 100 bps. In comparing two bonds, it is assumed that the "given change" is the same for both securities. When the government bond par curve is shifted up or down by the same amount to calculate effective duration and effective convexity, the events are described as "parallel" yield curve shifts. Because yield curves are rarely (if ever) straight lines, this shift may also be described as a "shape-preserving" shift to the yield curve. The key assumption is that all yields-to-maturity under consideration rise or fall by the same amount across the curve.

Although the assumption of a parallel shift in the yield curve is common in fixed-income analysis, it is not always realistic. The shape of the yield curve changes based on factors affecting the supply and demand of shorter-term versus longer-term securities. In fact, the term structure of bond yields (also called the "term structure of interest rates") is typically upward sloping. However, the **term structure of yield volatility** may have a different shape depending on a number of factors. The term structure of yield volatility is the relationship between the volatility of bond yields-to-maturity and times-to-maturity.

For example, a central bank engaging in expansionary monetary policy might cause the yield curve to steepen by reducing short-term interest rates. But this policy might cause greater *volatility* in short-term bond yields-to-maturity than in longer-term bonds, resulting in a downward-sloping term structure of yield volatility. Longer-term bond yields are mostly determined by future inflation and economic growth expectations. Those expectations often tend to be less volatile.

The importance of yield volatility in measuring interest rate risk is that bond price changes are products of two factors: (1) the impact *per* basis-point change in the yield-to-maturity and (2) the *number* of basis points in the yield-to-maturity change. The first factor is duration or the combination of duration and convexity, and the second factor is the yield volatility. For example, consider a 5-year bond with a modified duration of 4.5 and a 30-year bond with a modified duration of 18.0. Clearly, for a *given* change in yield-to-maturity, the 30-year bond represents much more interest rate risk to an investor who has a short-term horizon. In fact, the 30-year bond appears to have *four times* the risk given the ratio of the modified durations. But that assumption neglects the possibility that the 30-year bond might have half the yield volatility of the 5-year bond.

Equation 14, restated here, summarizes the two factors.

$$\%\Delta PV^{Full} \approx (-\text{AnnModDur} \times \Delta\text{Yield}) + \left[\frac{1}{2} \times \text{AnnConvexity} \times (\Delta\text{Yield})^2\right]$$

The estimated percentage change in the bond price depends on the modified duration and convexity as well as on the yield-to-maturity change. Parallel shifts between two bond yields and along a benchmark yield curve are common assumptions in fixed-income analysis. However, an analyst must be aware that non-parallel shifts frequently occur in practice.

EXAMPLE 15

A fixed-income analyst is asked to rank three bonds in terms of interest rate risk. Interest rate risk here means the potential price decrease on a percentage basis given a sudden change in financial market conditions. The increases in the yields-to-maturity represent the "worst case" for the scenario being considered.

Bond	Modified Duration	Convexity	ΔYield
A	3.72	12.1	25 bps
B	5.81	40.7	15 bps
C	12.39	158.0	10 bps

The modified duration and convexity statistics are annualized. ΔYield is the increase in the annual yield-to-maturity. Rank the bonds in terms of interest rate risk.

Solution:

Calculate the estimated percentage price change for each bond:

Bond A:

$$(-3.72 \times 0.0025) + \left[\frac{1}{2} \times 12.1 \times (0.0025)^2\right] = -0.009262$$

Bond B:

$$(-5.81 \times 0.0015) + \left[\frac{1}{2} \times 40.7 \times (0.0015)^2\right] = -0.008669$$

Bond C:

$$(-12.39 \times 0.0010) + \left[\frac{1}{2} \times 158.0 \times (0.0010)^2\right] = -0.012311$$

Based on these assumed changes in the yield-to-maturity and the modified duration and convexity risk measures, Bond C has the highest degree of interest rate risk (a potential loss of 1.2311%), followed by Bond A (a potential loss of 0.9262%) and Bond B (a potential loss of 0.8669%).

10.2 Investment Horizon, Macaulay Duration, and Interest Rate Risk

Although short-term interest rate risk is a concern to some investors, other investors have a long-term horizon. Day-to-day changes in bond prices cause *unrealized* capital gains and losses. Those unrealized gains and losses might need to be accounted for in financial statements. This section considers a long-term investor concerned only with the total return over the investment horizon. Therefore, interest rate risk is important to this investor. The investor faces coupon reinvestment risk as well as market price risk if the bond needs to be sold prior to maturity.

Earlier, we discussed examples of interest rate risk using a 10-year, 8% annual coupon payment bond that is priced at 85.503075 per 100 of par value. The bond's yield-to-maturity is 10.40%. A key result in Example 3 is that an investor with a 10-year time horizon is concerned only with coupon reinvestment risk. This situation assumes, of course, that the issuer makes all of the coupon and principal payments as scheduled. The buy-and-hold investor has a higher total return if interest rates rise (see Example 3) and a lower total return if rates fall (see Example 5). The investor in Examples 4 and 6 has a four-year horizon. This investor faces market price risk in addition to coupon reinvestment risk. In fact, the market price risk dominates because this investor has a higher total return if interest rates fall (see Example 6) and a lower return if rates rise (see Example 4).

Now, consider a third investor who has a seven-year time horizon. If interest rates remain at 10.40%, the future value of reinvested coupon interest is 76.835787 per 100 of par value.

$$\left[8 \times (1.1040)^6\right] + \left[8 \times (1.1040)^5\right] + \left[8 \times (1.1040)^4\right] + \left[8 \times (1.1040)^3\right] + \left[8 \times (1.1040)^2\right] + \left[8 \times (1.1040)^1\right] + 8 = 76.835787$$

The bond is sold for a price of 94.073336, assuming that the bond stays on the constant-yield price trajectory and continues to be "pulled to par."

$$\frac{8}{(1.1040)^1} + \frac{8}{(1.1040)^2} + \frac{108}{(1.1040)^3} = 94.073336$$

The total return is 170.909123 (= 76.835787 + 94.073336) per 100 of par value, and the horizon yield, as expected, is 10.40%.

$$85.503075 = \frac{170.909123}{(1+r)^7}, \quad r = 0.1040$$

Following Examples 3 and 4, assume that the yield-to-maturity on the bond rises to 11.40%. Also, coupon interest is now reinvested each year at 11.40%. The future value of reinvested coupons becomes 79.235183 per 100 of par value.

$$\left[8 \times (1.1140)^6\right] + \left[8 \times (1.1140)^5\right] + \left[8 \times (1.1140)^4\right] + \left[8 \times (1.1140)^3\right] + \left[8 \times (1.1140)^2\right] + \left[8 \times (1.1140)^1\right] + 8 = 79.235183$$

After receiving the seventh coupon payment, the bond is sold. There is a capital loss because the price, although much higher than at purchase, is below the constant-yield price trajectory.

$$\frac{8}{(1.1140)^1} + \frac{8}{(1.1140)^2} + \frac{108}{(1.1140)^3} = 91.748833$$

The total return is 170.984016 (= 79.235183 + 91.748833) per 100 of par value and the holding-period rate of return is 10.407%.

$$85.503075 = \frac{170.984016}{(1+r)^7}, \quad r = 0.10407$$

Following Examples 5 and 6, assume that the coupon reinvestment rates and the bond yield-to-maturity fall to 9.40%. The future value of reinvested coupons is 74.512177.

$$\left[8 + (1.0940)^6\right] + \left[8 + (1.0940)^5\right] + \left[8 + (1.0940)^4\right] + \left[8 + (1.0940)^3\right] +$$
$$\left[8 + (1.0940)^2\right] + \left[8 + (1.0940)^1\right] + 8 = 74.52177$$

The bond is sold at a capital gain because the price is above the constant-yield price trajectory.

$$\frac{8}{(1.0940)^1} + \frac{8}{(1.0940)^2} + \frac{108}{(1.0940)^3} = 96.481299$$

The total return is 170.993476 (= 74.512177 + 96.481299) per 100 of par value, and the horizon yield is 10.408%.

$$85.503075 = \frac{170.993476}{(1+r)^7}, \quad r = 0.10408$$

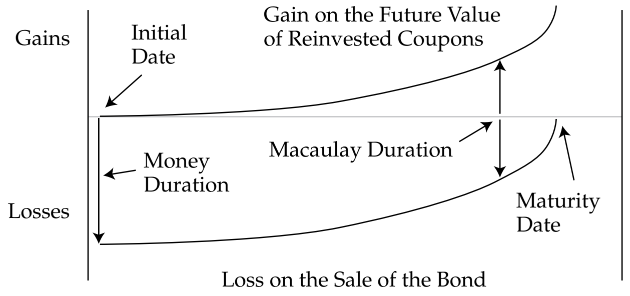
These results are summarized in the following table to reveal the remarkable outcome: The total returns and horizon yields are virtually the same. The investor with the 7-year horizon, unlike those having a 4- or 10-year horizon, achieves the same holding-period rate of return whether interest rates rise, fall, or remain the same. Note that the terms "horizon yield" and "holding-period rate of return" are used interchangeably in this reading. Sometimes "horizon yield" refers to yields on bonds that need to be sold at the end of the investor's holding period.

Interest Rate	Future Value of Reinvested Coupon	Sale Price	Total Return	Horizon Yield
9.40%	74.512177	96.481299	170.993476	10.408%
10.40%	76.835787	94.073336	170.909123	10.400%
11.40%	79.235183	91.748833	170.984016	10.407%

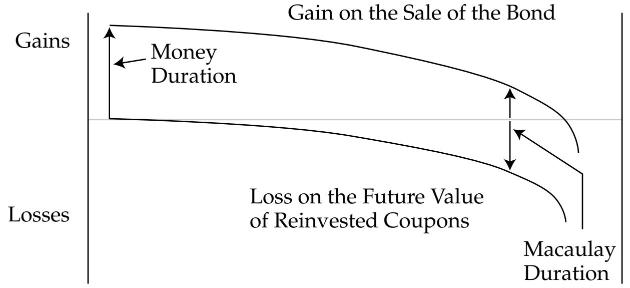
This particular bond was chosen as an example to demonstrate an important property of Macaulay duration: For a particular assumption about yield volatility, Macaulay duration indicates the investment horizon for which coupon reinvestment risk and market price risk offset each other. In Exhibit 2, the Macaulay duration of this 10-year, 8% annual payment bond is calculated to be 7.0029 years. This is one of the applications for duration in which "years" is meaningful and in which Macaulay duration is used rather than modified duration. The particular assumption about yield volatility is that there is a one-time "parallel" shift in the yield curve that occurs before the next coupon payment date. Exhibit 13 illustrates this property of bond duration, assuming that the bond is initially priced at par value.

Exhibit 13 Interest Rate Risk, Macaulay Duration, and the Investment Horizon

A. Interest Rates Rise



B. Interest Rates Fall



As demonstrated in Panel A of Exhibit 13, when interest rates rise, duration measures the immediate drop in value. In particular, the money duration indicates the change in price. Then as time passes, the bond price is "pulled to par." The gain in the future value of reinvested coupons starts small but builds over time as more coupons are received. The curve indicates the additional future value of reinvested coupons because of the higher interest rate. At some point in the lifetime of the bond, those two effects offset each other and the gain on reinvested coupons is equal to the loss on the sale of the bond. That point in time is the Macaulay duration statistic.

The same pattern is displayed in the Panel B when interest rates fall, which leads to a reduction in the bond yield and the coupon reinvestment rate. There is an immediate jump in the bond price, as measured by the money duration, but then the "pull to par" effect brings the price down as time passes. The impact from reinvesting at a lower rate starts small but then becomes more significant over time. The loss on reinvested coupons is with respect to the future value if interest rates had not fallen. Once again, the bond's Macaulay duration indicates the point in time when the two effects offset each other and the gain on the sale of the bond matches the loss on coupon reinvestment.

The earlier numerical example and Exhibit 13 allow for a statement of the general relationships among interest rate risk, the Macaulay duration, and the investment horizon.

1 When the investment horizon is greater than the Macaulay duration of a bond, coupon reinvestment risk dominates market price risk. The investor's risk is to lower interest rates.

2 When the investment horizon is equal to the Macaulay duration of a bond, coupon reinvestment risk offsets market price risk.

3 When the investment horizon is less than the Macaulay duration of the bond, market price risk dominates coupon reinvestment risk. The investor's risk is to higher interest rates.

In the numerical example, the Macaulay duration of the bond is 7.0 years. Statement 1 reflects the investor with the 10-year horizon; Statement 2, the investor with the 7-year horizon; and Statement 3, the investor with the 4-year horizon.

The difference between the Macaulay duration of a bond and the investment horizon is called the **duration gap**. The duration gap is a bond's Macaulay duration minus the investment horizon. The investor with the 10-year horizon has a negative duration gap and currently is at risk of lower rates. The investor with the 7-year horizon has a duration gap of zero and currently is hedged against interest rate risk. The investor with the 4-year horizon has a positive duration gap and currently is at risk of higher rates. The word "currently" is important because interest rate risk is connected to an *immediate* change in the bond's yield-to-maturity and the coupon reinvestment rates. As time passes, the investment horizon is reduced and the Macaulay duration of the bond also changes. Therefore, the duration gap changes as well.

EXAMPLE 16

An investor plans to retire in 10 years. As part of the retirement portfolio, the investor buys a newly issued, 12-year, 8% annual coupon payment bond. The bond is purchased at par value, so its yield-to-maturity is 8.00% stated as an effective annual rate.

1 Calculate the approximate Macaulay duration for the bond, using a 1 bp increase and decrease in the yield-to-maturity and calculating the new prices per 100 of par value to six decimal places.

2 Calculate the duration gap at the time of purchase.

3 Does this bond at purchase entail the risk of higher or lower interest rates? Interest rate risk here means an immediate, one-time, parallel yield curve shift.

Solution to 1:

The approximate modified duration of the bond is 7.5361. PV_0 = 100, PV_+ = 99.924678, and PV_- = 100.075400.

$$PV_+ = \frac{8}{(1.0801)^1} + \cdots + \frac{108}{(1.0801)^{12}} = 99.924678$$

$$PV_- = \frac{8}{(1.0799)^1} + \cdots + \frac{108}{(1.0799)^{12}} = 100.075400$$

$$\text{ApproxModDur} = \frac{100.075400 - 99.924678}{2 \times 0.0001 \times 100} = 7.5361$$

The approximate Macaulay duration is 8.1390 (= 7.5361 × 1.08).

Solution to 2:

Given an investment horizon of 10 years, the duration gap for this bond at purchase is negative: 8.1390 – 10 = –1.8610.

Solution to 3:

A negative duration gap entails the risk of lower interest rates. To be precise, the risk is an immediate, one-time, parallel, downward yield curve shift because coupon reinvestment risk dominates market price risk. The loss from reinvesting coupons at a rate lower than 8% is larger than the gain from selling the bond at a price above the constant-yield price trajectory.

CREDIT AND LIQUIDITY RISK

11

l. explain how changes in credit spread and liquidity affect yield-to-maturity of a bond and how duration and convexity can be used to estimate the price effect of the changes

The focus of this reading is to demonstrate how bond duration and convexity estimate the bond price change, either in percentage terms or in units of currency, given an assumed yield-to-maturity change. This section addresses the *source* of the change in the yield-to-maturity. In general, the yield-to-maturity on a corporate bond is composed of a government *benchmark* yield and a *spread* over that benchmark. A change in the bond's yield-to-maturity can originate in either component or a combination of the two.

The key point is that for a traditional (option-free) fixed-rate bond, the same duration and convexity statistics apply for a change in the benchmark yield as for a change in the spread. The "building blocks" approach covered in an earlier reading shows that these yield-to-maturity changes can be broken down further. A change in the benchmark yield can arise from a change in either the expected inflation rate or the expected real rate of interest. A change in the spread can arise from a change in the credit risk of the issuer or in the liquidity of the bond. Therefore, for a fixed-rate bond, the "inflation duration," the "real rate duration," the "credit duration," and the "liquidity duration" are all the same number. The inflation duration would indicate the change in the bond price if expected inflation were to change by a certain amount. In the same manner, the real rate duration would indicate the bond price change if the real rate were to go up or down. The credit duration and liquidity duration would indicate the price sensitivity that would arise from changes in those building blocks in the yield-to-maturity. A bond with a modified duration of 5.00 and a convexity of 32.00 will appreciate in value by about 1.26% if its yield-to-maturity goes down by 25 bps: $(-5.00 \times -0.0025) + [1/2 \times 32.00 \times (-0.0025)^2] = +0.0126$, regardless of the *source* of the yield-to-maturity change.

Suppose that the yield-to-maturity on a corporate bond is 6.00%. If the benchmark yield is 4.25%, the spread is 1.75%. An analyst believes that credit risk makes up 1.25% of the spread and liquidity risk, the remaining 0.50%. Credit risk includes the probability of default as well as the recovery of assets if default does occur. A credit rating downgrade or an adverse change in the ratings outlook for a borrower reflects a higher risk of default. Liquidity risk refers to the transaction costs associated with selling a bond. In general, a bond with greater frequency of trading and a higher volume of trading provides fixed-income investors with more opportunity to purchase or sell the security and thus has less liquidity risk. In practice, there is a difference between the *bid* (or purchase) and the *offer* (or sale) price. This difference depends on the type of bond, the size of the transaction, and the time of execution, among other factors. For instance, government bonds often trade at just a few basis points between the purchase and sale prices. More thinly traded corporate bonds can have a much wider difference between the bid and offer prices.

The problem for a fixed-income analyst is that it is rare for the changes in the components of the overall yield-to-maturity to occur in isolation. In practice, the analyst is concerned with the *interaction* between changes in benchmark yields and spreads, between changes in expected inflation and the expected real rate, and between changes in credit and liquidity risk. For example, during a financial crisis, a "flight to quality" can cause government benchmark yields to fall as credit spreads widen. An unexpected credit downgrade on a corporate bond can result in greater credit as well as liquidity risk.

EXAMPLE 17

The (flat) price on a fixed-rate corporate bond falls one day from 92.25 to 91.25 per 100 of par value because of poor earnings and an unexpected ratings downgrade of the issuer. The (annual) modified duration for the bond is 7.24. Which of the following is *closest* to the estimated change in the credit spread on the corporate bond, assuming benchmark yields are unchanged?

A 15 bps

B 100 bps

C 108 bps

Solution:

Given that the price falls from 92.25 to 91.25, the percentage price decrease is 1.084%.

$$\frac{91.25 - 92.25}{92.25} = -0.01084$$

Given an annual modified duration of 7.24, the change in the yield-to-maturity is 14.97 bps.

$$-0.01084 \approx -7.24 \times \Delta\text{Yield},\ \Delta\text{Yield} = 0.001497$$

Therefore, the answer is A. The change in price reflects a credit spread increase on the bond of about 15 bps.

12 EMPIRICAL DURATION

m describe the difference between empirical duration and analytical duration

The approach taken in this reading to estimate duration and convexity statistics using mathematical formulas is often referred to as **analytical duration**. These estimates of the impact of benchmark yield changes on bond prices implicitly assume that government bond yields and spreads are independent variables that are uncorrelated with one another. Analytical duration offers a reasonable approximation of the price–yield relationship in many situations, but fixed-income professionals often use historical data in statistical models that incorporate various factors affecting bond prices to calculate **empirical duration** estimates. These estimates calculated over time and in different interest rate environments inform the fixed-income portfolio decision-making process and will be addressed in detail in later readings.

For instance, in the "flight to quality" example cited earlier in which investors sell risky assets during market turmoil and purchase default-risk-free government bonds, we might expect analytical and empirical duration estimates to differ among bond types. For example, on the one hand, for a government bond with little or no credit risk, we would expect analytical and empirical duration to be similar because benchmark yield changes largely drive bond prices. On the other hand, the same macroeconomic factors driving government bond yields lower in a market stress scenario will cause high-yield bond credit spreads to widen because of an increase in expected default risk. Since credit spreads and benchmark yields are negatively correlated under this scenario, wider credit spreads will partially or fully offset the decline in government benchmark yields, resulting in lower empirical duration estimates than analytical duration estimates. Importantly, analysts must consider the correlation between benchmark yields and credit spreads when deciding whether to use empirical or analytical duration estimates.

EXAMPLE 18

AFC Investment Ltd. is a fixed-income investment firm that actively manages a government bond fund and a corporate bond fund. Holdings of the government bond fund are mainly medium-term US Treasury securities but also include debt of highly rated developed-market sovereign issuers. About half of the corporate bond fund is invested in investment-grade issues, and the other half consists of high-yield issues, all with a mix of maturities and from a mix of North American, European, and Asian companies.

Explain why empirical duration is likely to be a more accurate risk measure for AFC's corporate bond fund than for its government bond fund.

Solution:

The government bond fund includes debt securities of the US government and other highly rated developed-market sovereign issuers. Since benchmark yields are the primary driver of changes in overall bond yields in this fund, the results of analytical duration and empirical duration should be broadly similar.

The corporate bond fund includes a wide variety of debt securities with varying levels of credit quality and liquidity and, therefore, different credit and liquidity spreads. Interactions between benchmark yield changes and credit and liquidity spreads would tend to offset each other, particularly during stressed market conditions, making empirical duration significantly lower than analytical duration. As a result, empirical duration may be the more accurate risk measure for the corporate bond fund.

SUMMARY

This reading covers the risk and return characteristics of fixed-rate bonds. The focus is on the widely used measures of interest rate risk—duration and convexity. These statistics are used extensively in fixed-income analysis. The following are the main points made in the reading:

- The three sources of return on a fixed-rate bond purchased at par value are: (1) receipt of the promised coupon and principal payments on the scheduled dates, (2) reinvestment of coupon payments, and (3) potential capital gains, as well as losses, on the sale of the bond prior to maturity.

- For a bond purchased at a discount or premium, the rate of return also includes the effect of the price being "pulled to par" as maturity nears, assuming no default.
- The total return is the future value of reinvested coupon interest payments and the sale price (or redemption of principal if the bond is held to maturity).
- The horizon yield (or holding period rate of return) is the internal rate of return between the total return and purchase price of the bond.
- Coupon reinvestment risk increases with a higher coupon rate and a longer reinvestment time period.
- Capital gains and losses are measured from the carrying value of the bond and not from the purchase price. The carrying value includes the amortization of the discount or premium if the bond is purchased at a price below or above par value. The carrying value is any point on the constant-yield price trajectory.
- Interest income on a bond is the return associated with the passage of time. Capital gains and losses are the returns associated with a change in the value of a bond as indicated by a change in the yield-to-maturity.
- The two types of interest rate risk on a fixed-rate bond are coupon reinvestment risk and market price risk. These risks offset each other to a certain extent. An investor gains from higher rates on reinvested coupons but loses if the bond is sold at a capital loss because the price is below the constant-yield price trajectory. An investor loses from lower rates on reinvested coupon but gains if the bond is sold at a capital gain because the price is above the constant-yield price trajectory.
- Market price risk dominates coupon reinvestment risk when the investor has a short-term horizon (relative to the time-to-maturity on the bond).
- Coupon reinvestment risk dominates market price risk when the investor has a long-term horizon (relative to the time-to-maturity)—for instance, a buy-and-hold investor.
- Bond duration, in general, measures the sensitivity of the full price (including accrued interest) to a change in interest rates.
- Yield duration statistics measuring the sensitivity of a bond's full price to the bond's own yield-to-maturity include the Macaulay duration, modified duration, money duration, and price value of a basis point.
- Curve duration statistics measuring the sensitivity of a bond's full price to the benchmark yield curve are usually called "effective durations."
- Macaulay duration is the weighted average of the time to receipt of coupon interest and principal payments, in which the weights are the shares of the full price corresponding to each payment. This statistic is annualized by dividing by the periodicity (number of coupon payments or compounding periods in a year).
- Modified duration provides a linear estimate of the percentage price change for a bond given a change in its yield-to-maturity.
- Approximate modified duration approaches modified duration as the change in the yield-to-maturity approaches zero.
- Effective duration is very similar to approximate modified duration. The difference is that approximate modified duration is a yield duration statistic that measures interest rate risk in terms of a change in the bond's own yield-to-maturity, whereas effective duration is a curve duration statistic that measures interest rate risk assuming a parallel shift in the benchmark yield curve.

- Key rate duration is a measure of a bond's sensitivity to a change in the benchmark yield curve at specific maturity segments. Key rate durations can be used to measure a bond's sensitivity to changes in the shape of the yield curve.
- Bonds with an embedded option do not have a meaningful internal rate of return because future cash flows are contingent on interest rates. Therefore, effective duration is the appropriate interest rate risk measure, not modified duration.
- The effective duration of a traditional (option-free) fixed-rate bond is its sensitivity to the benchmark yield curve, which can differ from its sensitivity to its own yield-to-maturity. Therefore, modified duration and effective duration on a traditional (option-free) fixed-rate bond are not necessarily equal.
- During a coupon period, Macaulay and modified durations decline smoothly in a "saw-tooth" pattern, assuming the yield-to-maturity is constant. When the coupon payment is made, the durations jump upward.
- Macaulay and modified durations are inversely related to the coupon rate and the yield-to-maturity.
- Time-to-maturity and Macaulay and modified durations are *usually* positively related. They are *always* positively related on bonds priced at par or at a premium above par value. They are *usually* positively related on bonds priced at a discount below par value. The exception is on long-term, low-coupon bonds, on which it is possible to have a lower duration than on an otherwise comparable shorter-term bond.
- The presence of an embedded call option reduces a bond's effective duration compared with that of an otherwise comparable non-callable bond. The reduction in the effective duration is greater when interest rates are low and the issuer is more likely to exercise the call option.
- The presence of an embedded put option reduces a bond's effective duration compared with that of an otherwise comparable non-putable bond. The reduction in the effective duration is greater when interest rates are high and the investor is more likely to exercise the put option.
- The duration of a bond portfolio can be calculated in two ways: (1) the weighted average of the time to receipt of *aggregate* cash flows and (2) the weighted average of the durations of individual bonds that compose the portfolio.
- The first method to calculate portfolio duration is based on the cash flow yield, which is the internal rate of return on the aggregate cash flows. It cannot be used for bonds with embedded options or for floating-rate notes.
- The second method is simpler to use and quite accurate when the yield curve is relatively flat. Its main limitation is that it assumes a parallel shift in the yield curve in that the yields on all bonds in the portfolio change by the same amount.
- Money duration is a measure of the price change in terms of units of the currency in which the bond is denominated.
- The price value of a basis point (PVBP) is an estimate of the change in the full price of a bond given a 1 bp change in the yield-to-maturity.
- Modified duration is the primary, or first-order, effect on a bond's percentage price change given a change in the yield-to-maturity. Convexity is the secondary, or second-order, effect. It indicates the change in the modified duration as the yield-to-maturity changes.
- Money convexity is convexity times the full price of the bond. Combined with money duration, money convexity estimates the change in the full price of a bond in units of currency given a change in the yield-to-maturity.

- Convexity is a positive attribute for a bond. Other things being equal, a more convex bond appreciates in price more than a less convex bond when yields fall and depreciates less when yields rise.
- Effective convexity is the second-order effect on a bond price given a change in the benchmark yield curve. It is similar to approximate convexity. The difference is that approximate convexity is based on a yield-to-maturity change and effective convexity is based on a benchmark yield curve change.
- Callable bonds have negative effective convexity when interest rates are low. The increase in price when the benchmark yield is reduced is less in absolute value than the decrease in price when the benchmark yield is raised.
- The change in a bond price is the product of: (1) the impact per basis-point change in the yield-to-maturity and (2) the number of basis points in the yield change. The first factor is estimated by duration and convexity. The second factor depends on yield volatility.
- The investment horizon is essential in measuring the interest rate risk on a fixed-rate bond.
- For a particular assumption about yield volatility, the Macaulay duration indicates the investment horizon for which coupon reinvestment risk and market price risk offset each other. The assumption is a one-time parallel shift to the yield curve in which the yield-to-maturity and coupon reinvestment rates change by the same amount in the same direction.
- When the investment horizon is greater than the Macaulay duration of the bond, coupon reinvestment risk dominates price risk. The investor's risk is to lower interest rates. The duration gap is negative.
- When the investment horizon is equal to the Macaulay duration of the bond, coupon reinvestment risk offsets price risk. The duration gap is zero.
- When the investment horizon is less than the Macaulay duration of the bond, price risk dominates coupon reinvestment risk. The investor's risk is to higher interest rates. The duration gap is positive.
- Credit risk involves the probability of default and degree of recovery if default occurs, whereas liquidity risk refers to the transaction costs associated with selling a bond.
- For a traditional (option-free) fixed-rate bond, the same duration and convexity statistics apply if a change occurs in the benchmark yield or a change occurs in the spread. The change in the spread can result from a change in credit risk or liquidity risk.
- In practice, there often is interaction between changes in benchmark yields and in the spread over the benchmark.
- Empirical duration uses statistical methods and historical bond prices to derive the price–yield relationship for specific bonds or bond portfolios.

REFERENCES

Smith, Donald J. 2014. *Bond Math: The Theory behind the Formulas*. 2nd ed. Hoboken, NJ: John Wiley & Sons.

PRACTICE PROBLEMS

1 A "buy-and-hold" investor purchases a fixed-rate bond at a discount and holds the security until it matures. Which of the following sources of return is *least likely* to contribute to the investor's total return over the investment horizon, assuming all payments are made as scheduled?

- A Capital gain
- B Principal payment
- C Reinvestment of coupon payments

2 Which of the following sources of return is *most likely* exposed to interest rate risk for an investor of a fixed-rate bond who holds the bond until maturity?

- A Capital gain or loss
- B Redemption of principal
- C Reinvestment of coupon payments

3 An investor purchases a bond at a price above par value. Two years later, the investor sells the bond. The resulting capital gain or loss is measured by comparing the price at which the bond is sold to the:

- A carrying value.
- B original purchase price.
- C original purchase price value plus the amortized amount of the premium.

The following information relates to Questions 4–6

An investor purchases a nine-year, 7% annual coupon payment bond at a price equal to par value. After the bond is purchased and before the first coupon is received, interest rates increase to 8%. The investor sells the bond after five years. Assume that interest rates remain unchanged at 8% over the five-year holding period.

4 Per 100 of par value, the future value of the reinvested coupon payments at the end of the holding period is *closest* to:

- A 35.00.
- B 40.26.
- C 41.07.

5 The capital gain/loss per 100 of par value resulting from the sale of the bond at the end of the five-year holding period is *closest* to a:

- A loss of 8.45.
- B loss of 3.31.
- C gain of 2.75.

6 Assuming that all coupons are reinvested over the holding period, the investor's five-year horizon yield is *closest* to:

- A 5.66%.

B 6.62%.

C 7.12%.

---

7 An investor buys a three-year bond with a 5% coupon rate paid annually. The bond, with a yield-to-maturity of 3%, is purchased at a price of 105.657223 per 100 of par value. Assuming a 5-basis point change in yield-to-maturity, the bond's approximate modified duration is *closest* to:

A 2.78.

B 2.86.

C 5.56.

8 Which of the following statements about duration is correct? A bond's:

A effective duration is a measure of yield duration.

B modified duration is a measure of curve duration.

C modified duration cannot be larger than its Macaulay duration (assuming a positive yield-to-maturity).

9 An investor buys a 6% annual payment bond with three years to maturity. The bond has a yield-to-maturity of 8% and is currently priced at 94.845806 per 100 of par. The bond's Macaulay duration is *closest* to:

A 2.62.

B 2.78.

C 2.83.

10 The interest rate risk of a fixed-rate bond with an embedded call option is *best* measured by:

A effective duration.

B modified duration.

C Macaulay duration.

11 Which of the following is *most* appropriate for measuring a bond's sensitivity to shaping risk?

A Key rate duration

B Effective duration

C Modified duration

12 A Canadian pension fund manager seeks to measure the sensitivity of her pension liabilities to market interest rate changes. The manager determines the present value of the liabilities under three interest rate scenarios: a base rate of 7%, a 100 basis point increase in rates up to 8%, and a 100 basis point drop in rates down to 6%. The results of the manager's analysis are presented below:

Interest Rate Assumption	Present Value of Liabilities
6%	CAD510.1 million
7%	CAD455.4 million
8%	CAD373.6 million

The effective duration of the pension fund's liabilities is *closest* to:

A 1.49.

B 14.99.

C 29.97.

13 Which of the following statements about Macaulay duration is correct?

A A bond's coupon rate and Macaulay duration are positively related.

B A bond's Macaulay duration is inversely related to its yield-to-maturity.

C The Macaulay duration of a zero-coupon bond is less than its time-to-maturity.

14 Assuming no change in the credit risk of a bond, the presence of an embedded put option:

A reduces the effective duration of the bond.

B increases the effective duration of the bond.

C does not change the effective duration of the bond.

15 A bond portfolio consists of the following three fixed-rate bonds. Assume annual coupon payments and no accrued interest on the bonds. Prices are per 100 of par value.

Bond	Maturity	Market Value	Price	Coupon	Yield-to-Maturity	Modified Duration
A	6 years	170,000	85.0000	2.00%	4.95%	5.42
B	10 years	120,000	80.0000	2.40%	4.99%	8.44
C	15 years	100,000	100.0000	5.00%	5.00%	10.38

The bond portfolio's modified duration is *closest* to:

A 7.62.

B 8.08.

C 8.20.

16 A limitation of calculating a bond portfolio's duration as the weighted average of the yield durations of the individual bonds that compose the portfolio is that it:

A assumes a parallel shift to the yield curve.

B is less accurate when the yield curve is less steeply sloped.

C is not applicable to portfolios that have bonds with embedded options.

17 Using the information below, which bond has the *greatest* money duration per 100 of par value assuming annual coupon payments and no accrued interest?

Bond	Time-to-Maturity	Price Per 100 of Par Value	Coupon Rate	Yield-to-Maturity	Modified Duration
A	6 years	85.00	2.00%	4.95%	5.42
B	10 years	80.00	2.40%	4.99%	8.44
C	9 years	85.78	3.00%	5.00%	7.54

A Bond A

B Bond B

C Bond C

18 A bond with exactly nine years remaining until maturity offers a 3% coupon rate with annual coupons. The bond, with a yield-to-maturity of 5%, is priced at 85.784357 per 100 of par value. The estimated price value of a basis point for the bond is *closest* to:

A 0.0086.

B 0.0648.

C 0.1295.

19 The "second-order" effect on a bond's percentage price change given a change in yield-to-maturity can be *best* described as:

A duration.

B convexity.

C yield volatility.

20 A bond is currently trading for 98.722 per 100 of par value. If the bond's yield-to-maturity (YTM) rises by 10 basis points, the bond's full price is expected to fall to 98.669. If the bond's YTM decreases by 10 basis points, the bond's full price is expected to increase to 98.782. The bond's approximate convexity is *closest* to:

A 0.071.

B 70.906.

C 1,144.628.

21 A bond has an annual modified duration of 7.020 and annual convexity of 65.180. If the bond's yield-to-maturity decreases by 25 basis points, the expected percentage price change is *closest* to:

A 1.73%.

B 1.76%.

C 1.78%.

22 A bond has an annual modified duration of 7.140 and annual convexity of 66.200. The bond's yield-to-maturity is expected to increase by 50 basis points. The expected percentage price change is *closest* to:

A –3.40%.

B –3.49%.

C –3.57%.

23 Which of the following statements relating to yield volatility is *most* accurate? If the term structure of yield volatility is downward sloping, then:

A short-term rates are higher than long-term rates.

B long-term yields are more stable than short-term yields.

C short-term bonds will always experience greater price fluctuation than long-term bonds.

24 The holding period for a bond at which the coupon reinvestment risk offsets the market price risk is *best* approximated by:

A duration gap.

B modified duration.

C Macaulay duration.

25 When the investor's investment horizon is less than the Macaulay duration of the bond she owns:

A the investor is hedged against interest rate risk.

B reinvestment risk dominates, and the investor is at risk of lower rates.

C market price risk dominates, and the investor is at risk of higher rates.

26 An investor purchases an annual coupon bond with a 6% coupon rate and exactly 20 years remaining until maturity at a price equal to par value. The investor's investment horizon is eight years. The approximate modified duration of the bond is 11.470 years. The duration gap at the time of purchase is *closest* to:

A –7.842.

B 3.470.

C 4.158.

27 A manufacturing company receives a ratings upgrade and the price increases on its fixed-rate bond. The price increase was *most likely* caused by a(n):

A decrease in the bond's credit spread.

B increase in the bond's liquidity spread.

C increase of the bond's underlying benchmark rate.

28 Empirical duration is likely the best measure of the impact of yield changes on portfolio value, especially under stressed market conditions, for a portfolio consisting of:

A 100% sovereign bonds of several AAA rated euro area issuers.

B 100% covered bonds of several AAA rated euro area corporate issuers.

C 25% AAA rated sovereign bonds, 25% AAA rated corporate bonds, and 50% high-yield (i.e., speculative-grade) corporate bonds, all from various euro area sovereign and corporate issuers.

SOLUTIONS

1 A is correct. A capital gain is least likely to contribute to the investor's total return. There is no capital gain (or loss) because the bond is held to maturity. The carrying value of the bond at maturity is par value, the same as the redemption amount. When a fixed-rate bond is held to its maturity, the investor receives the principal payment at maturity. This principal payment is a source of return for the investor. A fixed-rate bond pays periodic coupon payments, and the reinvestment of these coupon payments is a source of return for the investor. The investor's total return is the redemption of principal at maturity and the sum of the reinvested coupons.

2 C is correct. Because the fixed-rate bond is held to maturity (a "buy-and-hold" investor), interest rate risk arises entirely from changes in coupon reinvestment rates. Higher interest rates increase income from reinvestment of coupon payments, and lower rates decrease income from coupon reinvestment. There will not be a capital gain or loss because the bond is held until maturity. The carrying value at the maturity date is par value, the same as the redemption amount. The redemption of principal does not expose the investor to interest rate risk. The risk to a bond's principal is credit risk.

3 A is correct. Capital gains (losses) arise if a bond is sold at a price above (below) its constant-yield price trajectory. A point on the trajectory represents the carrying value of the bond at that time. That is, the capital gain/loss is measured from the bond's carrying value, the point on the constant-yield price trajectory, and not from the original purchase price. The carrying value is the original purchase price plus the amortized amount of the discount if the bond is purchased at a price below par value. If the bond is purchased at a price above par value, the carrying value is the original purchase price minus (not plus) the amortized amount of the premium. The amortized amount for each year is the change in the price between two points on the trajectory.

4 C is correct. The future value of reinvested cash flows at 8% after five years is closest to 41.07 per 100 of par value.

$$\left[7 \times (1.08)^4\right] + \left[7 \times (1.08)^3\right] + \left[7 \times (1.08)^2\right] + \left[7 \times (1.08)^1\right] + 7 = 41.0662$$

The 6.07 difference between the sum of the coupon payments over the five-year holding period (35) and the future value of the reinvested coupons (41.07) represents the "interest-on-interest" gain from compounding.

5 B is correct. The capital loss is closest to 3.31 per 100 of par value. After five years, the bond has four years remaining until maturity and the sale price of the bond is 96.69, calculated as

$$\frac{7}{(1.08)^1} + \frac{7}{(1.08)^2} + \frac{7}{(1.08)^3} + \frac{107}{(1.08)^4} = 96.69$$

The investor purchased the bond at a price equal to par value (100). Because the bond was purchased at a price equal to its par value, the carrying value is par value. Therefore, the investor experienced a capital loss of 96.69 – 100 = –3.31.

6 B is correct. The investor's five-year horizon yield is closest to 6.62%. After five years, the sale price of the bond is 96.69 (from problem 5) and the future value of reinvested cash flows at 8% is 41.0662 (from problem 4) per 100 of par value. The total return is 137.76 (= 41.07 + 96.69), resulting in a realized five-year horizon yield of 6.62%:

$$100.00 = \frac{137.76}{(1+r)^5}, \quad r = 0.0662$$

7 A is correct. The bond's approximate modified duration is closest to 2.78. Approximate modified duration is calculated as

$$\text{ApproxModDur} = \frac{(PV_-) - (PV_+)}{2 \times (\Delta\text{Yield}) \times (PV_0)}$$

Lower yield-to-maturity by 5 bps to 2.95%:

$$PV_- = \frac{5}{(1+0.0295)^1} + \frac{5}{(1+0.0295)^2} + \frac{5+100}{(1+0.0295)^3} = 105.804232$$

Increase yield-to-maturity by 5 bps to 3.05%:

$$PV_+ = \frac{5}{(1+0.0305)^1} + \frac{5}{(1+0.0305)^2} + \frac{5+100}{(1+0.0305)^3} = 105.510494$$

$PV_0 = 105.657223$, $\Delta\text{Yield} = 0.0005$

$$\text{ApproxModDur} = \frac{105.804232 - 105.510494}{2 \times 0.0005 \times 105.657223} = 2.78$$

8 C is correct. A bond's modified duration cannot be larger than its Macaulay duration assuming a positive yield-to-maturity. The formula for modified duration is

$$\text{ModDur} = \frac{\text{MacDur}}{1+r}$$

where r is the bond's yield-to-maturity per period. Therefore, ModDur will typically be less than MacDur.

Effective duration is a measure of curve duration. Modified duration is a measure of yield duration.

9 C is correct. The bond's Macaulay duration is closest to 2.83. Macaulay duration (MacDur) is a weighted average of the times to the receipt of cash flow. The weights are the shares of the full price corresponding to each coupon and principal payment.

Period	Cash Flow	Present Value	Weight	Period × Weight
1	6	5.555556	0.058575	0.058575
2	6	5.144033	0.054236	0.108472
3	106	84.146218	0.887190	2.661570
		94.845806	1.000000	2.828617

Thus, the bond's Macaulay duration (MacDur) is 2.83.

Alternatively, Macaulay duration can be calculated using the following closed-form formula:

$$\text{MacDur} = \left\{\frac{1+r}{r} - \frac{1+r+\left[N \times (c-r)\right]}{c \times \left[(1+r)^N - 1\right] + r}\right\} - (t/T)$$

$$\text{MacDur} = \left\{\frac{1.08}{0.08} - \frac{1.08+\left[3 \times (0.06-0.08)\right]}{0.06 \times \left[(1.08)^3 - 1\right] + 0.08}\right\} - 0$$

MacDur = 13.50 – 10.67 = 2.83

10 A is correct. The interest rate risk of a fixed-rate bond with an embedded call option is best measured by effective duration. A callable bond's future cash flows are uncertain because they are contingent on future interest rates. The issuer's decision to call the bond depends on future interest rates. Therefore, the yield-to-maturity on a callable bond is not well defined. Only effective duration, which takes into consideration the value of the call option, is the appropriate interest rate risk measure. Yield durations like Macaulay and modified durations are not relevant for a callable bond because they assume no changes in cash flows when interest rates change.

11 A is correct. Key rate duration is used to measure a bond's sensitivity to a shift at one or more maturity segments of the yield curve which result in a change to yield curve shape. Modified and effective duration measure a bond's sensitivity to parallel shifts in the entire curve.

12 B is correct. The effective duration of the pension fund's liabilities is closest to 14.99. The effective duration is calculated as follows:

$$\text{EffDur} = \frac{(PV_-) - (PV_+)}{2 \times (\Delta\text{Curve}) \times (PV_0)}$$

PV_0 = 455.4, PV_+ = 373.6, PV_- = 510.1, and ΔCurve = 0.0100.

$$\text{EffDur} = \frac{510.1 - 373.6}{2 \times 0.0100 \times 455.4} = 14.99$$

13 B is correct. A bond's yield-to-maturity is inversely related to its Macaulay duration: The higher the yield-to-maturity, the lower its Macaulay duration and the lower the interest rate risk. A higher yield-to-maturity decreases the weighted average of the times to the receipt of cash flow, and thus decreases the Macaulay duration.

A bond's coupon rate is inversely related to its Macaulay duration: The lower the coupon, the greater the weight of the payment of principal at maturity. This results in a higher Macaulay duration. Zero-coupon bonds do not pay periodic coupon payments; therefore, the Macaulay duration of a zero-coupon bond is its time-to-maturity.

14 A is correct. The presence of an embedded put option reduces the effective duration of the bond, especially when rates are rising. If interest rates are low compared with the coupon rate, the value of the put option is low and the impact of the change in the benchmark yield on the bond's price is very similar to the impact on the price of a non-putable bond. But when benchmark interest rates rise, the put option becomes more valuable to the investor. The ability to sell the bond at par value limits the price depreciation as rates rise. The presence of an embedded put option reduces the sensitivity of the bond price to changes in the benchmark yield, assuming no change in credit risk.

15 A is correct. The portfolio's modified duration is closest to 7.62. Portfolio duration is commonly estimated as the market-value-weighted average of the yield durations of the individual bonds that compose the portfolio.

The total market value of the bond portfolio is 170,000 + 120,000 + 100,000 = 390,000.

The portfolio duration is 5.42 × (170,000/390,000) + 8.44 × (120,000/390,000) + 10.38 × (100,000/390,000) = 7.62.

16 A is correct. A limitation of calculating a bond portfolio's duration as the weighted average of the yield durations of the individual bonds is that this measure implicitly assumes a parallel shift to the yield curve (all rates change by the same amount in the same direction). In reality, interest rate changes frequently result in a steeper or flatter yield curve. This approximation of the "theoretically correct" portfolio duration is *more* accurate when the yield curve is flatter (less steeply sloped). An advantage of this approach is that it can be used with portfolios that include bonds with embedded options. Bonds with embedded options can be included in the weighted average using the effective durations for these securities.

17 B is correct. Bond B has the greatest money duration per 100 of par value. Money duration (MoneyDur) is calculated as the annual modified duration (AnnModDur) times the full price (PV^{Full}) of the bond including accrued interest. Bond B has the highest money duration per 100 of par value.

MoneyDur = AnnModDur × PV^{Full}

MoneyDur of Bond A = 5.42 × 85.00 = 460.70

MoneyDur of Bond B = 8.44 × 80.00 = 675.20

MoneyDur of Bond C = 7.54 × 85.78 = 646.78

18 B is correct. The PVBP is closest to 0.0648. The formula for the price value of a basis point is

$$\text{PVBP} = \frac{(PV_-) - (PV_+)}{2}$$

where

PVBP = price value of a basis point
PV_- = full price calculated by lowering the yield-to-maturity by 1 bp
PV_+ = full price calculated by raising the yield-to-maturity by 1 bp

Lowering the yield-to-maturity by 1 bp to 4.99% results in a bond price of 85.849134:

$$PV_- = \frac{3}{(1 + 0.0499)^1} + \cdots + \frac{3 + 100}{(1 + 0.0499)^9} = 85.849134$$

Increasing the yield-to-maturity by 1 bp to 5.01% results in a bond price of 85.719638:

$$PV_+ = \frac{3}{(1 + 0.0501)^1} + \cdots + \frac{3 + 100}{(1 + 0.0501)^9} = 85.719638$$

$$\text{PVBP} = \frac{85.849134 - 85.719638}{2} = 0.06475$$

Alternatively, the PVBP can be derived using modified duration:

$$\text{ApproxModDur} = \frac{(PV_-) - (PV_+)}{2 \times (\Delta\text{Yield}) \times (PV_0)}$$

$$\text{ApproxModDur} = \frac{85.849134 - 85.719638}{2 \times 0.0001 \times 85.784357} = 7.548$$

PVBP = 7.548 × 85.784357 × 0.0001 = 0.06475

19 B is correct. Convexity measures the "second order" effect on a bond's percentage price change given a change in yield-to-maturity. Convexity adjusts the percentage price change estimate provided by modified duration to better approximate the true relationship between a bond's price and its yield-to-maturity which is a curved line (convex).

Duration estimates the change in the bond's price along the straight line that is tangent to this curved line ("first order" effect). Yield volatility measures the magnitude of changes in the yields along the yield curve.

20 B is correct. The bond's approximate convexity is closest to 70.906. Approximate convexity (ApproxCon) is calculated using the following formula:

$$\text{ApproxCon} = [PV_- + PV_+ - (2 \times PV_0)]/(\Delta\text{Yield}^2 \times PV_0)$$

where

PV_- = new price when the yield-to-maturity is decreased
PV_+ = new price when the yield-to-maturity is increased
PV_0 = original price
ΔYield = change in yield-to-maturity
ApproxCon = $[98.782 + 98.669 - (2 \times 98.722)]/(0.001^2 \times 98.722) = 70.906$

21 C is correct. The expected percentage price change is closest to 1.78%. The convexity-adjusted percentage price change for a bond given a change in the yield-to-maturity is estimated by

$$\%\Delta PV^{Full} \approx [-\text{AnnModDur} \times \Delta\text{Yield}] + [0.5 \times \text{AnnConvexity} \times (\Delta\text{Yield})^2]$$

$$\%\Delta PV^{Full} \approx [-7.020 \times (-0.0025)] + [0.5 \times 65.180 \times (-0.0025)^2] = 0.017754,$$ or 1.78%

22 B is correct. The expected percentage price change is closest to –3.49%. The convexity-adjusted percentage price change for a bond given a change in the yield-to-maturity is estimated by

$$\%\Delta PV^{Full} \approx [-\text{AnnModDur} \times \Delta\text{Yield}] + [0.5 \times \text{AnnConvexity} \times (\Delta\text{Yield})^2]$$

$$\%\Delta PV^{Full} \approx [-7.140 \times 0.005] + [0.5 \times 66.200 \times (0.005)^2] = -0.034873,$$ or –3.49%

23 B is correct. If the term structure of yield volatility is downward-sloping, then short-term bond yields-to-maturity have greater volatility than for long-term bonds. Therefore, long-term yields are more stable than short-term yields. Higher volatility in short-term rates does not necessarily mean that the level of short-term rates is higher than long-term rates. With a downward-sloping term structure of yield volatility, short-term bonds will not always experience greater price fluctuation than long-term bonds. The estimated percentage change in a bond price depends on the modified duration and convexity as well as on the yield-to-maturity change.

24 C is correct. When the holder of a bond experiences a one-time parallel shift in the yield curve, the Macaulay duration statistic identifies the number of years necessary to hold the bond so that the losses (or gains) from coupon reinvestment offset the gains (or losses) from market price changes. The duration gap is the difference between the Macaulay duration and the investment horizon. Modified duration approximates the percentage price change of a bond given a change in its yield-to-maturity.

25 C is correct. The duration gap is equal to the bond's Macaulay duration minus the investment horizon. In this case, the duration gap is positive, and price risk dominates coupon reinvestment risk. The investor risk is to higher rates.

The investor is hedged against interest rate risk if the duration gap is zero; that is, the investor's investment horizon is equal to the bond's Macaulay duration. The investor is at risk of lower rates only if the duration gap is negative; that is, the investor's investment horizon is greater than the bond's Macaulay duration. In this case, coupon reinvestment risk dominates market price risk.

26 C is correct. The duration gap is closest to 4.158. The duration gap is a bond's Macaulay duration minus the investment horizon. The approximate Macaulay duration is the approximate modified duration times one plus the yield-to-maturity. It is 12.158 (= 11.470 × 1.06).

Given an investment horizon of eight years, the duration gap for this bond at purchase is positive: 12.158 – 8 = 4.158. When the investment horizon is less than the Macaulay duration of the bond, the duration gap is positive, and price risk dominates coupon reinvestment risk.

27 A is correct. The price increase was most likely caused by a decrease in the bond's credit spread. The ratings upgrade most likely reflects a lower expected probability of default and/or a greater level of recovery of assets if default occurs. The decrease in credit risk results in a smaller credit spread. The increase in the bond price reflects a decrease in the yield-to-maturity due to a smaller credit spread. The change in the bond price was not due to a change in liquidity risk or an increase in the benchmark rate.

28 C is correct. Empirical duration is the best measure—better than analytical duration—of the impact of yield changes on portfolio value, especially under stressed market conditions, for a portfolio consisting of a variety of different bonds from different issuers, such as the portfolio described in Answer C. In this portfolio, credit spread changes on the high-yield bonds may partly or fully offset yield changes on the AAA rated sovereign bonds and spread changes on the AAA rated corporate bonds; this interaction is best captured using empirical duration. The portfolios described in Answers A and B consist of the same types of bonds from similar issuers—sovereign bonds from similar-rated sovereign issuers (A) and covered bonds from similar-rated corporate issuers (B)—so empirical and analytical durations should be roughly similar in each of these portfolios.

READING

44

Fundamentals of Credit Analysis

by Christopher L. Gootkind, CFA

Christopher L. Gootkind, CFA, is at Loomis, Sayles & Company, LP (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** describe credit risk and credit-related risks affecting corporate bonds;
☐	**b.** describe default probability and loss severity as components of credit risk;
☐	**c.** describe seniority rankings of corporate debt and explain the potential violation of the priority of claims in a bankruptcy proceeding;
☐	**d.** compare and contrast corporate issuer credit ratings and issue credit ratings and describe the rating agency practice of "notching";
☐	**e.** explain risks in relying on ratings from credit rating agencies;
☐	**f.** explain the four Cs (Capacity, Collateral, Covenants, and Character) of traditional credit analysis;
☐	**g.** calculate and interpret financial ratios used in credit analysis;
☐	**h.** evaluate the credit quality of a corporate bond issuer and a bond of that issuer, given key financial ratios of the issuer and the industry;
☐	**i.** describe macroeconomic, market, and issuer-specific factors that influence the level and volatility of yield spreads;
☐	**j.** explain special considerations when evaluating the credit of high-yield, sovereign, and non-sovereign government debt issuers and issues.

The author would like to thank several of his Fixed Income Research colleagues at Loomis, Sayles & Company for their assistance with this reading: Paul Batterton, CFA, Diana Moskowitz, CFA, Diana Monteith, Shannon O'Mara, CFA, and Laura Sarlo, CFA.

1 INTRODUCTION

a describe credit risk and credit-related risks affecting corporate bonds

b describe default probability and loss severity as components of credit risk

With bonds outstanding worth many trillions of US dollars, the debt markets play a critical role in the global economy. Companies and governments raise capital in the debt market to fund current operations; buy equipment; build factories, roads, bridges, airports, and hospitals; acquire assets; and so on. By channeling savings into productive investments, the debt markets facilitate economic growth. Credit analysis has a crucial function in the debt capital markets—efficiently allocating capital by properly assessing credit risk, pricing it accordingly, and repricing it as risks change. How do fixed-income investors determine the riskiness of that debt, and how do they decide what they need to earn as compensation for that risk?

In the sections that follow, we cover basic principles of credit analysis, which may be broadly defined as the process by which credit risk is evaluated. Readers will be introduced to the definition of credit risk, the interpretation of credit ratings, the four Cs of traditional credit analysis, and key financial measures and ratios used in credit analysis. We explain, among other things, how to compare bond issuer creditworthiness within a given industry as well as across industries and how credit risk is priced in the bond market.

Our coverage focuses primarily on analysis of corporate debt; however, credit analysis of sovereign and nonsovereign, particularly municipal, government bonds will also be addressed. Structured finance, a segment of the debt markets that includes securities backed by such pools of assets as residential and commercial mortgages as well as other consumer loans, will not be covered here.

We first introduce the key components of credit risk—default probability and loss severity— along with such credit-related risks as spread risk, credit migration risk, and liquidity risk. We then discuss the relationship between credit risk and the capital structure of the firm before turning attention to the role of credit rating agencies. We also explore the process of analyzing the credit risk of corporations and examine the impact of credit spreads on risk and return. Finally, we look at special considerations applicable to the analysis of (i) high-yield (low-quality) corporate bonds and (ii) government bonds.

2 CREDIT RISK

Credit risk is the risk of loss resulting from the borrower (issuer of debt) failing to make full and timely payments of interest and/or principal. Credit risk has two components. The first is known as **default risk**, or **default probability**, which is the probability that a borrower defaults—that is, fails to meet its obligation to make full and timely payments of principal and interest according to the terms of the debt security. The second component is **loss severity** (also known as "loss given default") in the event of default—that is, the portion of a bond's value (including unpaid interest) an investor loses. A default can lead to losses of various magnitudes. In most instances, in the event of default, bondholders will recover some value, so there will not be a total loss on the investment. Thus, credit risk is reflected in the distribution of potential losses that may arise if the investor is not paid in full and on time. Although it is sometimes important to consider the entire distribution of potential losses and their respective probabilities—for instance, when losses have a disproportionate impact on the various tranches of a securitized pool of loans—it is often convenient to summarize the risk with a single default probability and loss severity and focus on the **expected loss**:

Expected loss = Default probability × Loss severity given default.

The loss severity, and hence the expected loss, can be expressed as either a monetary amount (e.g., €450,000) or as a percentage of the principal amount (e.g., 45%). The latter form of expression is generally more useful for analysis because it is independent of the amount of investment. Loss severity is often expressed as (1 – Recovery rate), where the recovery rate is the percentage of the principal amount recovered in the event of default.

Because default risk (default probability) is quite low for most high-quality debt issuers, bond investors tend to focus primarily on assessing this probability and devote less effort to assessing the potential loss severity arising from default. However, as an issuer's default risk rises, investors will focus more on what the recovery rate might be in the event of default. This issue will be discussed in more detail later. Important credit-related risks include the following:

- **Spread risk**. Corporate bonds and other "credit-risky" debt instruments typically trade at a yield premium, or spread, to bonds that have been considered "default-risk free," such as US Treasury bonds or German government bonds. Yield spreads, expressed in basis points, widen based on factors specific to the issuer, such as a decline in creditworthiness, sometimes referred to as credit migration or downgrade risk, or factors associated with the market as a whole, such as an increase in **market liquidity risk** or a general aversion to risk during periods of financial distress.
- **Credit migration risk** or **downgrade risk**. This is the risk that a bond issuer's creditworthiness deteriorates, or migrates lower, leading investors to believe the risk of default is higher and thus causing the yield spreads on the issuer's bonds to widen and the price of its bonds to fall. The term "downgrade" refers to action by the major bond rating agencies, whose role will be covered in more detail later.
- **Market liquidity risk**. This is the risk that the price at which investors can actually transact—buying or selling—may differ from the price indicated in the market. To compensate investors for the risk that there may not be sufficient market liquidity for them to buy or sell bonds in the quantity they desire, the spread or yield premium on corporate bonds includes a market liquidity component, in addition to a credit risk component. Unlike stocks, which trade on exchanges, most markets bonds trade primarily over the counter through broker–dealers trading for their own accounts. Their ability and willingness to make markets, as reflected in the bid–ask spread, is an important determinant of market liquidity risk. The two main issuer-specific factors that affect market liquidity risk are (1) the size of the issuer (that is, the amount of publicly traded debt an issuer has outstanding) and (2) the credit quality of the issuer. In general, the less debt an issuer has outstanding, the less frequently its debt trades and thus the higher the market liquidity risk. And the lower the quality of the issuer, the higher the market liquidity risk.

During times of financial stress or crisis, such as in late 2008, market liquidity can decline sharply, causing yield spreads on corporate bonds and other credit-risky debt to widen and their prices to drop.

EXAMPLE 1

Defining Credit Risk

1 Which of the following *best* defines credit risk?

   A The probability of default times the severity of loss given default

   B The loss of principal and interest payments in the event of bankruptcy

   C The risk of not receiving full interest and principal payments on a timely basis

2 Which of the following is the *best* measure of credit risk?

   A The expected loss

   B The severity of loss

   C The probability of default

3 Which of the following is NOT credit or credit-related risk?

   A Default risk

   B Interest rate risk

   C Downgrade or credit migration risk

Solution to 1:

C is correct. Credit risk is the risk that the borrower will not make full and timely payments.

Solution to 2:

A is correct. The expected loss captures both of the key components of credit risk: (the product of) the probability of default and the loss severity in the event of default. Neither component alone fully reflects the risk.

Solution to 3:

B is correct. Bond price changes due to general interest rate movements are not considered credit risk.

3 CAPITAL STRUCTURE, SENIORITY RANKING, AND RECOVERY RATES

c describe seniority rankings of corporate debt and explain the potential violation of the priority of claims in a bankruptcy proceeding

The various debt obligations of a given borrower will not necessarily all have the same **seniority ranking**, or priority of payment. In this section, we will introduce the topic of an issuer's capital structure and discuss the various types of debt claims that may arise from that structure, as well as their ranking and how those rankings can influence recovery rates in the event of default. The term "creditors" is used throughout our coverage to mean holders of debt instruments, such as bonds and bank loans. Unless specifically stated, it does not include such obligations as trade credit, tax liens, or employment-related obligations.

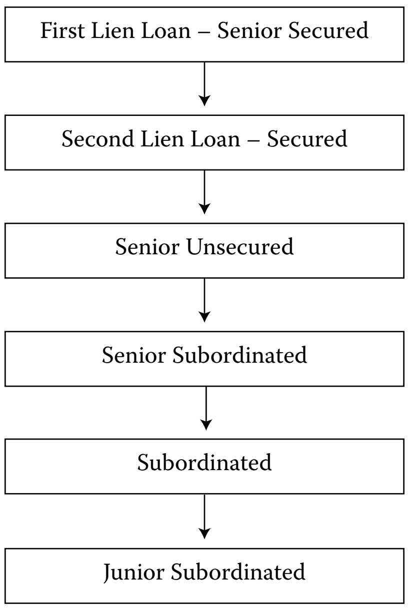
3.1 Capital Structure

The composition and distribution across operating units of a company's debt and equity—including bank debt, bonds of all seniority rankings, preferred stock, and common equity—is referred to as its **capital structure**. Some companies and industries have straightforward capital structures, with all the debt equally ranked and issued by one main operating entity. Other companies and industries, due to their frequent acquisitions and divestitures (e.g., media companies or conglomerates) or high levels of regulation (e.g., banks and utilities), tend to have more complicated capital structures. Companies in these industries often have many different subsidiaries, or operating companies, that have their own debt outstanding and parent holding companies that also issue debt, with different levels or rankings of seniority. Similarly, the cross-border operations of multinational corporations tend to increase the complexity of their capital structures.

3.2 Seniority Ranking

Just as borrowers can issue debt with many different maturity dates and coupons, they can also have many different rankings in terms of seniority. The ranking refers to the priority of payment, with the most senior or highest-ranking debt having the first claim on the cash flows and assets of the issuer. This level of seniority can affect the value of an investor's claim in the event of default and restructuring. Broadly, there is **secured debt** and **unsecured debt**. Secured debt means the debtholder has a direct claim—a pledge from the issuer—on certain assets and their associated cash flows. Unsecured bondholders have only a general claim on an issuer's assets and cash flows. In the event of default, unsecured debtholders' claims rank below (i.e., get paid after) those of secured creditors under what's known as the **priority of claims**.

Exhibit 1 Seniority Ranking



Within each category of debt are finer gradations of types and rankings. Within secured debt, there is first mortgage and first lien debt, which are the highest-ranked debt in terms of priority of repayment. **First mortgage debt** or loan refers to the pledge of a specific property (e.g., a power plant for a utility or a specific casino for a gaming company). **First lien debt** or loan refers to a pledge of certain assets that could include buildings but might also include property and equipment, licenses, patents, brands, and so on. There can also be **second lien**, or even third lien, secured debt, which, as the name implies, has a secured interest in the pledged assets but ranks below first lien debt in both collateral protection and priority of payment.

Within unsecured debt, there can also be finer gradations and seniority rankings. The highest-ranked unsecured debt is senior unsecured debt. It is the most common type of all corporate bonds outstanding. Other, lower-ranked debt includes **subordinated debt** and junior subordinated debt. Among the various creditor classes, these obligations have among the lowest priority of claims and frequently have little or no recovery in the event of default. That is, their loss severity can be as high as 100%. (See Exhibit 1 for a sample seniority ranking.) For regulatory and capital purposes, banks in Europe and the United States have issued debt and debt-like securities that rank even lower than subordinated debt, typically referred to as hybrids or trust preferred, and are intended to provide a capital cushion in times of financial distress. Many of them did not work as intended during the global financial crisis that began in 2008, and most were phased out, potentially to be replaced by more effective instruments that automatically convert to equity in certain circumstances.

Companies issue—and investors buy—debt with different seniority rankings for many reasons. Issuers are interested in optimizing their cost of capital—finding the right mix of the various types of both debt and equity—for their industry and type of business. Issuers may offer secured debt because that is what the market (i.e., investors) may require—given a company's perceived riskiness or because secured debt is generally lower cost due to the reduced credit risk inherent in its higher priority of claims. Or, issuers may offer subordinated debt because (1) they believe it is less expensive than issuing equity, as debtholders require a lower rate of return due to their superior place in line in the event of default; (2) doing so prevents dilution to existing shareholders; (3) it is typically less restrictive than issuing senior debt; and (4) investors are willing to buy it because they believe the yield being offered is adequate compensation for the risk they perceive.

EXAMPLE 2

Seniority Ranking

The Acme Company has senior unsecured bonds as well as both first and second lien debt in its capital structure. Which ranks higher with respect to priority of claims: senior unsecured bonds or second lien debt?

Solution:

Second lien debt ranks higher than senior unsecured bonds because of its secured position.

3.3 Recovery Rates

All creditors at the same level of the capital structure are treated as one class; thus, a senior unsecured bondholder whose debt is due in 30 years has the same pro rata claim in bankruptcy as one whose debt matures in six months. This provision is referred to as bonds ranking **pari passu** ("on an equal footing") in right of payment.

Defaulted debt will often continue to be traded by investors and broker–dealers based on their assessment that either in liquidation of the bankrupt company's assets or in reorganization, the bonds will have some recovery value. In the case of reorganization or restructuring (whether through formal bankruptcy or on a voluntary basis), new debt, equity, cash, or some combination thereof could be issued in exchange for the original defaulted debt.

As discussed, recovery rates vary by seniority of ranking in a company's capital structure, under the priority of claims treatment in bankruptcy. Over many decades, there have been enough defaults to generate statistically meaningful historical data on recovery rates by seniority ranking. Exhibit 2 provides recovery rates by seniority ranking for North American non-financial companies. For example, as shown in Exhibit 2, investors on average recovered 46.9% of the value of senior secured debt that defaulted in 2016 but only 29.2% of the value of senior unsecured issues that defaulted that year.

Exhibit 2 Average Corporate Debt Recovery Rates Measured by Ultimate Recoveries

	Emergence Year*			Default Year		
Seniority Ranking	**2017**	**2016**	**1987–2017**	**2017**	**2016**	**1987–2017**
Bank loans	81.3%	72.6%	80.4%	80.2%	78.3%	80.4%
Senior secured bonds	52.3%	35.9%	62.3%	57.5%	46.9%	62.3%
Senior unsecured bonds	54.1%	11.7%	47.9%	47.4%	29.2%	47.9%
Subordinated bonds	4.5%	6.6%	28.0%	NA	8.0%	28.0%

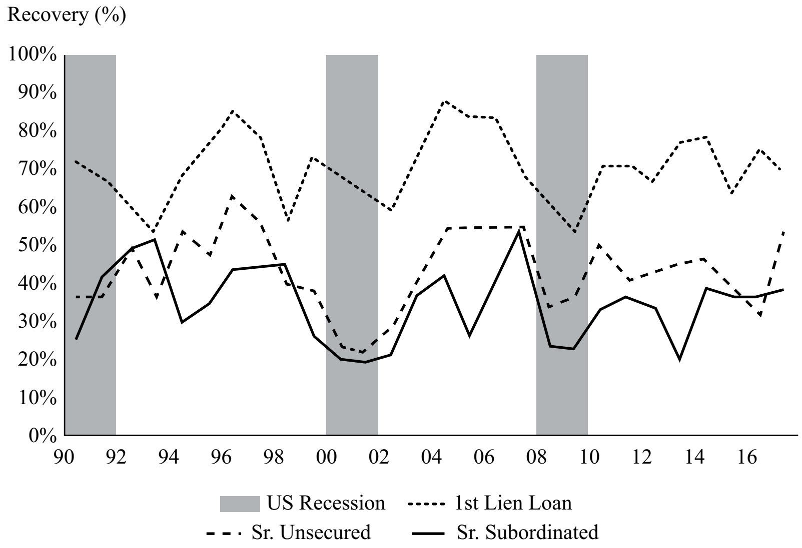
NA = not available.
*Emergence year is typically the year the defaulted company emerges from bankruptcy. Default year data refer to the recovery rate of debt that defaulted in that year (i.e., 2016 and 2017) or range of years (i.e., 1987–2017). Data are for North American nonfinancial companies.
Source: Moody's Investors Service's Ultimate Recovery Database.

A few things are worth noting:

1 **Recovery rates can vary widely by industry**. Companies that go bankrupt in industries that are in secular decline (e.g., newspaper publishing) will most likely have lower recovery rates than those that go bankrupt in industries merely suffering from a cyclical economic downturn.

2 **Recovery rates can also vary depending on when they occur in a credit cycle**. Credit cycles describe the changing availability—and pricing—of credit. When the economy is strong or improving, the willingness of lenders to extend credit on favorable terms is high. Conversely, when the economy is weak or weakening, lenders pull back, or "tighten" credit, by making it less available and more expensive. As shown in Exhibit 3, at or near the bottom of a credit cycle—which is almost always closely linked with an economic cycle—recoveries will tend to be lower than at other times in the credit cycle.

Exhibit 3 Global Recovery Rates by Seniority Ranking, 1990–2017



Source: Based on data from Moody's Investors Service's Ultimate Recovery Database.

3 **These recovery rates are averages**. In fact, there can be large variability, both across industries, as noted, as well as across companies within a given industry. Factors might include composition and proportion of debt across an issuer's capital structure. An abundance of secured debt will lead to smaller recovery rates on lower-ranked debt.

Understanding recovery rates is important because they are a key component of credit analysis and risk. Recall that the best measure of credit risk is expected loss—that is, probability of default times loss severity given default. And loss severity equals (1 – Recovery rate). Having an idea how much one can lose in the event of default is a critical factor in valuing credit, particularly lower-quality credit, as the default risk rises.

Priority of claims: Not always absolute. The priority of claims in bankruptcy—the idea that the highest-ranked creditors get paid out first, followed by the next level, and on down, like a waterfall—is well established and is often described as "absolute." In principle, in the event of bankruptcy or liquidation:

- Creditors with a secured claim have the right to the value of that specific property before any other claim. If the value of the pledged property is less than the amount of the claim, then the difference becomes a senior unsecured claim.
- Unsecured creditors have a right to be paid in full before holders of equity interests (common and preferred shareholders) receive value on their interests.
- Senior unsecured creditors take priority over all subordinated creditors. A creditor is senior unsecured unless expressly subordinated.

In practice, however, creditors with lower seniority and even shareholders may receive some consideration without more senior creditors being paid in full. Why might this be the case? In bankruptcy, there are different classes of claimants, and all classes that are impaired (that is, receive less than full claim) get to vote to confirm the plan of reorganization. This vote is subject to the absolute priority of claims. Either by consent of the various parties or by the judge's order, however, absolute priority may not be strictly enforced in the final plan. There may be disputes over the value of various assets in the bankruptcy estate (e.g., what is a plant, or a patent portfolio, worth?) or the present value or timing of payouts. For example, what is the value of the new debt I'm receiving for my old debt of a reorganized company before it emerges from bankruptcy?

Resolution of these disputes takes time, and cases can drag on for months and years. In the meantime, during bankruptcy, substantial expenses are being incurred for legal and accounting fees, and the value of the company may be declining as key employees leave, customers go elsewhere, and so on. Thus, to avoid the time, expense, and uncertainty over disputed issues, such as the value of property in the estate and the legality of certain claims, the various claimants have an incentive to negotiate and compromise. This frequently leads to creditors with lower seniority and other claimants (e.g., even shareholders) receiving more consideration than they are legally entitled to.

It's worth noting that in the United States, the bias is toward reorganization and recovery of companies in bankruptcy, whereas in other jurisdictions, such as the United Kingdom, the bias is toward liquidation of companies in bankruptcy and maximizing value to the banks and other senior creditors. It's also worth noting that bankruptcy and bankruptcy laws are very complex and can vary greatly by country, so it is difficult to generalize about how creditors will fare. As shown in the earlier chart, there is huge variability in recovery rates for defaulted debt. Every case is different.

EXAMPLE 3

Priority of Claims

1. Under which circumstance is a subordinated bondholder *most likely* to recover some value in a bankruptcy without a senior creditor getting paid in full? When:
   - **A** absolute priority rules are enforced.
   - **B** the various classes of claimants agree to it.
   - **C** the company is liquidated rather than reorganized.
2. In the event of bankruptcy, claims at the same level of the capital structure are:
   - **A** on an equal footing, regardless of size, maturity, or time outstanding.
   - **B** paid in the order of maturity from shortest to longest, regardless of size or time outstanding.
   - **C** paid on a first-in, first-out (FIFO) basis so that the longest-standing claims are satisfied first, regardless of size or maturity.

Solution to 1:

B is correct. All impaired classes get to vote on the reorganization plan. Negotiation and compromise are often preferable to incurring huge legal and accounting fees in a protracted bankruptcy process that would otherwise reduce the value of the estate for all claimants. This process may allow junior creditors (e.g., subordinated bondholders) to recover some value even though more senior creditors do not get paid in full.

Solution to 2:

A is correct. All claims at the same level of the capital structure are pari passu (on an equal footing).

4 RATING AGENCIES, CREDIT RATINGS, AND THEIR ROLE IN THE DEBT MARKETS

d compare and contrast corporate issuer credit ratings and issue credit ratings and describe the rating agency practice of "notching"

The major credit rating agencies—Moody's Investors Service ("Moody's"), Standard & Poor's ("S&P"), and Fitch Ratings ("Fitch")—play a central, if somewhat controversial, role in the credit markets. For the vast majority of outstanding bonds, at least two of the agencies provide ratings: a symbol-based measure of the potential risk of default of a particular bond or issuer of debt. In the public and quasi-public bond markets, underwritten by investment banks as opposed to privately placed, issuers won't offer, and investors won't buy, bonds that do not carry ratings from Moody's, S&P, or Fitch. This practice applies for all types of bonds: government or sovereign; entities with implicit or explicit guarantees from the government, such as Ginnie Mae in the United States and *Pfandbriefe* in Germany; supranational entities, such as the World Bank, which are owned by several governments; corporate; non-sovereign government; and mortgage- and asset-backed debt. How did the rating agencies attain such a dominant position in the credit markets? What are credit ratings, and what do they mean? How does the market use credit ratings? What are the risks of relying solely or excessively on credit ratings?

The history of the major rating agencies goes back more than 100 years. John Moody began publishing credit analysis and opinions on US railroads in 1909. S&P published its first ratings in 1916. They have grown in size and prominence since then. Many bond investors like the fact that there are independent analysts who meet with the issuer and often have access to material, non-public information, such as financial projections that investors cannot receive, to aid in the analysis. What has also proven very attractive to investors is that credit ratings provide direct and easy comparability of the relative credit riskiness of all bond issuers, within and across industries and bond types, although there is some debate about ratings comparability across the types of bonds. For instance, investigations conducted after the 2008–2009 global financial crisis suggested that, for a given rating category, municipal bonds have experienced a lower historical incidence of default than corporate debt.

Several factors have led to the near universal use of credit ratings in the bond markets and the dominant role of the major credit rating agencies. These factors include the following:

- Independent assessment of credit risk
- Ease of comparison across bond issuers, issues, and market segments
- Regulatory and statutory reliance and usage
- Issuer payment for ratings
- Huge growth of debt markets
- Development and expansion of bond portfolio management and the accompanying bond indexes

However, in the aftermath of the global financial crisis of 2008–2009, when the rating agencies were blamed for contributing to the crisis with their overly optimistic ratings on securities backed by subprime mortgages, attempts were made to reduce the role and dominant positions of the major credit rating agencies. New rules, regulations, and legislation were passed to require the agencies to be more transparent, reduce conflicts of interest, and stimulate more competition. The "issuer pay" model allows the distribution of ratings to a broad universe of investors and undoubtedly facilitated widespread reliance on ratings. Challenging the dominance of Moody's, S&P, and Fitch, additional credit rating agencies have emerged. Some credit rating agencies that are well-established in their home markets but are not so well known globally, such as Dominion Bond Rating Service (DBRS) in Canada and Japan Credit Rating Agency (JCR) in Japan, have tried to raise their profiles. The market dominance of the biggest credit rating agencies, however, remains largely intact.

4.1 Credit Ratings

The three major global credit rating agencies—Moody's, S&P, and Fitch—use similar, symbol-based ratings that are basically an assessment of a bond issue's risk of default. Exhibit 4 shows their long-term ratings ranked from highest to lowest. Ratings on short-term debt, although available, are not shown here.

Exhibit 4 Long-Term Rating Matrix: Investment Grade vs. Non-Investment Grade

		Moody's	S&P	Fitch
Investment Grade	High-Quality Grade	Aaa	AAA	AAA
		Aa1	AA+	AA+
		Aa2	AA	AA
		Aa3	AA–	AA–
	Upper-Medium Grade	A1	A+	A+
		A2	A	A
		A3	A–	A–
	Low-Medium Grade	Baa1	BBB+	BBB+
		Baa2	BBB	BBB
		Baa3	BBB–	BBB–
Non-Investment Grade ("Junk" or "High Yield")	Low Grade or Speculative Grade	Ba1	BB+	BB+
		Ba2	BB	BB
		Ba3	BB–	BB–
		B1	B+	B+
		B2	B	B
		B3	B–	B–
		Caa1	CCC+	CCC+
		Caa2	CCC	CCC
		Caa3	CCC–	CCC–
		Ca	CC	CC
		C	C	C
	Default	C	D	D

Bonds rated triple-A (Aaa or AAA) are said to be "of the highest quality, subject to the lowest level of credit risk" (Moody's Investors Service, see "Rating Symbols and Definitions", https://www.moodys.com/Pages/amr002002.aspx) and thus have extremely low probabilities of default. Double-A (Aa or AA) rated bonds are referred to as "high-quality grade" and are also regarded as having very low default risk. Bonds rated single-A are referred to as "upper-medium grade." Baa (Moody's) or BBB (S&P

and Fitch) are called “low-medium grade.” Bonds rated Baa3/BBB– or higher are called “investment grade.” Bonds rated Ba1 or lower by Moody’s and BB+ or lower by S&P and Fitch, respectively, have speculative credit characteristics and increasingly higher default risk. As a group, these bonds are referred to in a variety of ways: “low grade,” “speculative grade,” “non-investment grade,” “below investment grade,” “high yield,” and, in an attempt to reflect the extreme level of risk, some observers refer to these bonds as “junk bonds.” The D rating is reserved for securities that are already in default in S&P’s and Fitch’s scales. For Moody’s, bonds rated C are likely, but not necessarily, in default. Generally, issuers of bonds rated investment grade are more consistently able to access the debt markets and can borrow at lower interest rates than those rated below investment grade.

In addition, rating agencies will typically provide outlooks on their respective ratings—positive, stable, or negative—and may provide other indicators on the potential direction of their ratings under certain circumstances, such as “On Review for Downgrade” or “On CreditWatch for an Upgrade.” It should also be noted that, in support of the ratings they publish, the rating agencies also provide extensive written commentary and financial analysis on the obligors they rate, as well as summary industry statistics.

4.2 Issuer vs. Issue Ratings

Rating agencies will typically provide both issuer and issue ratings, particularly as they relate to corporate debt. Terminology used to distinguish between issuer and issue ratings includes corporate family rating (CFR) and corporate credit rating (CCR) or issuer credit rating and issue credit rating. An issuer credit rating is meant to address an obligor’s overall creditworthiness—its ability and willingness to make timely payments of interest and principal on its debt. The issuer credit rating usually applies to its senior unsecured debt.

Issue ratings refer to specific financial obligations of an issuer and take into consideration such factors as ranking in the capital structure (e.g., secured or subordinated). Although **cross-default provisions**, whereby events of default (such as non-payment of interest) on one bond trigger default on all outstanding debt, imply the same default probability for all issues, specific issues may be assigned different credit ratings—higher or lower—due to a rating adjustment methodology known as **notching**.

Notching

For the rating agencies, likelihood of default—default risk—is the primary factor in assigning their ratings. However, there are secondary factors as well. These factors include the priority of payment in the event of a default (e.g., secured versus senior unsecured versus subordinated) as well as potential loss severity in the event of default. Another factor considered by rating agencies is **structural subordination**, which can arise when a corporation with a holding company structure has debt at both its parent holding company and operating subsidiaries. Debt at the operating subsidiaries will get serviced by the cash flow and assets of the subsidiaries before funds can be passed (“upstreamed”) to the holding company to service debt at that level.

Recognizing these different payment priorities, and thus the potential for higher (or lower) loss severity in the event of default, the rating agencies have adopted a notching process whereby their credit ratings on issues can be moved up or down from the issuer rating, which is usually the rating applied to its senior unsecured debt. As a general rule, the higher the senior unsecured rating, the smaller the notching adjustment. The reason behind this is that the higher the rating, the lower the perceived risk of default; so, the need to “notch” the rating to capture the potential difference in loss severity is greatly reduced. For lower-rated credits, however, the risk of default is greater and thus the potential difference in loss from a lower (or higher) priority ranking is a bigger consideration in assessing an issue's credit riskiness. Thus, the rating agencies will typically apply larger rating adjustments. For example, S&P applies the following notching guidelines:

> A key principle is that investment-grade ratings focus more on timeliness, while non-investment-grade ratings give additional weight to recovery. For example, subordinated debt can be rated up to two notches below a non-investment-grade corporate credit rating, but one notch at most if the corporate credit rating is investment grade. Conversely, … the 'AAA' rating category need not be notched at all, while at the 'CCC' level the gaps may widen.
>
> The rationale for this convention is straightforward: as default risk increases, the concern over what can be recovered takes on greater relevance and, therefore, greater rating significance. Accordingly, the ultimate recovery aspect of ratings is given more weight as one moves down the rating spectrum.(Standard & Poor's, "Rating the Issue," in *Corporate Ratings Criteria 2008* [New York: Standard & Poor's, 2008]: 65)

Exhibit 5 is an example of S&P's notching criteria, as applied to Infor Software Parent LLC, Inc. (Infor). Infor is a US-based global software and services company whose corporate credit rating from S&P is B–. Note how the company's senior secured bonds are rated B, whereas its senior unsecured bonds are rated two notches lower at CCC+ and its holding company debt is rated even one notch lower at CCC.

Exhibit 5 Infor S&P Ratings Detail (as of December 2018)

Corporate credit rating	B–/Stable
Senior secured (3 issues)	B
Senior unsecured (2 issues)	CCC+
Holding company debt (1 issue)	CCC

Source: Standard & Poor's Financial Services, LLC.

4.3 ESG Ratings

ESG investing has received a great deal of attention and acceptance as a viable investment strategy, even as a source of alpha. ESG investing involves asset selection and investment decision making based on disciplined evaluation of one or all of these considerations:

- Environmental themes, such as investing in companies that are responding to consumer demand for sustainable practices, the reduction of carbon emissions, and other environmental agendas.
- Social themes, such as investing in companies committed to a diverse and inclusive workplace, minimizing salary gap, and community involvement programs.
- Governance themes, such as investing in companies committed to diverse board composition and strong oversight.

MSCI Inc., a global provider of market indexes and analytic tools, has launched a set of ratings that aim to measure a company's attitudes, practices, and advances related to ESG. Their rules-based methodology aims to identify and track leaders and laggards in the space. Companies are evaluated according to their exposure to ESG risks and how well they manage those risks relative to peers. The MSCI ESG rating scale has seven tiers and spans AAA to CCC. From best to worst, leaders receive

AAA or AA; the next echelon constitutes an average rating of A, BBB, or BB; while laggards receive B or CCC. MSCI has also added countries, mutual funds, and ETFs to their ESG rating system. Traditional rating agencies, such as Moody's and S&P, are evaluating ESG for borrowers under these criteria too.

ESG investing allows investors to align their investment objectives with long-term trends considered important to the overall sustainability of a business. Some evidence suggests that funds that value 'ethics,' by construction, also outperform key market indexes.

4.4 Risks in Relying on Agency Ratings

e explain risks in relying on ratings from credit rating agencies

The dominant position of the rating agencies in the global debt markets, and the near-universal use of their credit ratings on debt securities, suggests that investors believe they do a good job assessing credit risk. In fact, with a few exceptions (e.g., too high ratings on US subprime mortgage-backed securities issued in the mid-2000s, which turned out to be much riskier than expected), their ratings have proved quite accurate as a relative measure of default risk. For example, Exhibit 6 shows historical S&P one-year global corporate default rates by rating category for the 20-year period from 1998 to 2017. It measures the percentage of issues that defaulted in a given calendar year based on how they were rated at the beginning of the year.

Exhibit 6 Global Corporate Annual Default Rates by Rating Category (%)

	AAA	AA	A	BBB	BB	B	CCC/C
1998	0.00	0.00	0.00	0.41	0.82	4.63	42.86
1999	0.00	0.17	0.18	0.20	0.95	7.29	33.33
2000	0.00	0.00	0.27	0.37	1.16	7.70	35.96
2001	0.00	0.00	0.27	0.34	2.96	11.53	45.45
2002	0.00	0.00	0.00	1.01	2.89	8.21	44.44
2003	0.00	0.00	0.00	0.23	0.58	4.07	32.73
2004	0.00	0.00	0.08	0.00	0.44	1.45	16.18
2005	0.00	0.00	0.00	0.07	0.31	1.74	9.09
2006	0.00	0.00	0.00	0.00	0.30	0.82	13.33
2007	0.00	0.00	0.00	0.00	0.20	0.25	15.24
2008	0.00	0.38	0.39	0.49	0.81	4.09	27.27
2009	0.00	0.00	0.22	0.55	0.75	10.94	49.46
2010	0.00	0.00	0.00	0.00	0.58	0.86	22.62
2011	0.00	0.00	0.00	0.07	0.00	1.67	16.30
2012	0.00	0.00	0.00	0.00	0.03	1.57	27.52
2013	0.00	0.00	0.00	0.00	0.10	1.64	24.50
2014	0.00	0.00	0.00	0.00	0.00	0.78	17.42
2015	0.00	0.00	0.00	0.00	0.16	2.40	26.51
2016	0.00	0.00	0.00	0.00	0.47	3.70	33.17
2017	0.00	0.00	0.00	0.00	0.08	0.98	26.23
Mean	0.00	0.03	0.07	0.19	0.69	3.82	27.98

Exhibit 6 (Continued)

	AAA	AA	A	BBB	BB	B	CCC/C
Max	0.00	0.38	0.39	1.01	2.96	11.53	45.45
Min	0.00	0.00	0.00	0.00	0.00	0.25	9.09

Source: Based on data from Standard & Poor's Financial Services, LLC.

As Exhibit 6 shows, the highest-rated bonds have extremely low default rates. With very few exceptions, the lower the rating, the higher the annual rate of default, with bonds rated CCC and lower experiencing the highest default rates by far.

Relying on credit rating agency ratings, however, also has limitations and risks, including the following:

- **Credit ratings can change over time**. Over a long time period (e.g., many years), credit ratings can migrate—move up or down—significantly from what they were at the time of bond issuance. Using Standard & Poor's data, Exhibit 7 shows the average three-year migration (or "transition") by rating from 1981 to 2017. Note that the higher the credit rating, the greater the ratings stability. Even for AAA rated credits, however, only about 65% of the time did ratings remain in that rating category over a three-year period. (Of course, AAA rated credits can have their ratings move in only one direction—down.) A very small fraction of AAA rated credits became non-investment grade or defaulted within three years. For single-B rated credits, only 41% of the time did ratings remain in that rating category over three-year periods. This observation about how credit ratings can change over time isn't meant to be a criticism of the rating agencies. It is meant to demonstrate that creditworthiness can and does change—up or down—and that bond investors should not assume an issuer's credit rating will remain the same from time of purchase through the entire holding period.

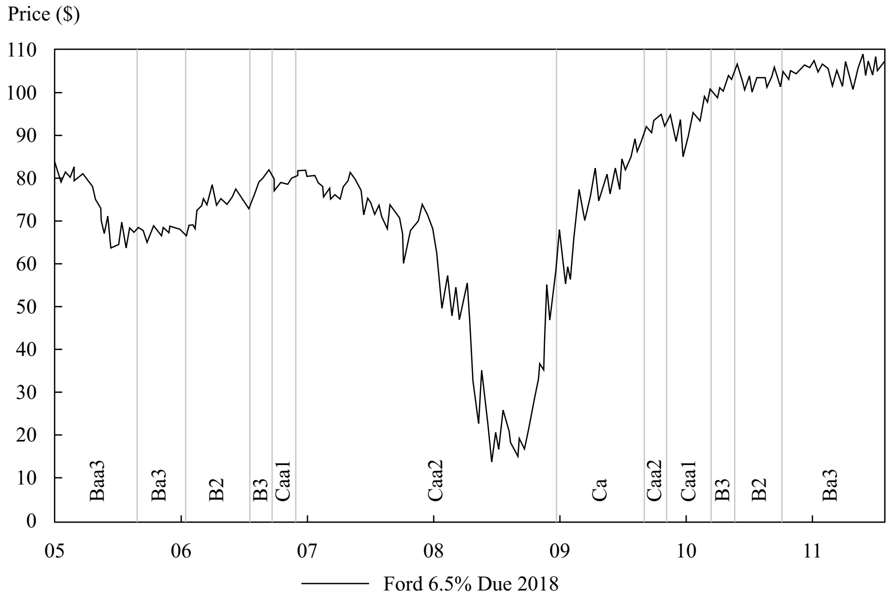
Exhibit 7 Average Three-Year Global Corporate Transition Rates, 1981–2017 (%)

From/To	AAA	AA	A	BBB	BB	B	CCC/C	D	NR
AAA	65.48	22.09	2.35	0.32	0.19	0.08	0.11	0.13	9.24
AA	1.21	66.14	18.53	2.06	0.35	0.22	0.03	0.12	11.33
A	0.06	4.07	68.85	11.72	1.30	0.44	0.09	0.25	13.21
BBB	0.02	0.28	8.42	64.66	7.11	1.64	0.30	0.87	16.70
BB	0.01	0.06	0.51	11.08	47.04	11.58	1.25	3.96	24.51
B	0.00	0.03	0.21	0.78	10.23	41.46	4.67	12.57	30.05
CCC/C	0.00	0.00	0.14	0.61	1.63	16.86	10.54	40.65	29.57

Notes: D = default. NR = not rated—that is, certain corporate issuers were no longer rated by S&P. This could occur for a variety of reasons, including issuers paying off their debt and no longer needing ratings.
Source: S&P Global Ratings, "2017 Annual Global Corporate Default Study and Rating Transitions" (5 April 2018): 53.

- **Credit ratings tend to lag the market's pricing of credit risk**. Bond prices and credit spreads frequently move more quickly because of changes in perceived creditworthiness than rating agencies change their ratings (or even outlooks) up or down. Bond prices and relative valuations can move every day, whereas bond ratings, appropriately, don't change that often. Even over long time periods, however, credit ratings can badly lag changes in bond prices. Exhibit 8 shows the price and Moody's rating of a bond from US automaker Ford Motor Company before, during, and after the global financial crisis in 2008. Note how the bond's price moved down sharply well before Moody's downgraded its credit rating—multiple times—and also how the bond's price began to recover—and kept recovering—well before Moody's upgraded its credit rating on Ford debt.

Exhibit 8 Historical Example: Ford Motor Company Senior Unsecured Debt: Price vs. Moody's Rating 2005–2011



Sources: Data based on Bloomberg Finance LP and Moody's Investors Service, Inc.

Moreover, particularly for certain speculative-grade credits, two bonds with similar ratings may trade at very different valuations. This is partly a result of the fact that credit ratings primarily try to assess the risk of default, whereas for low-quality credits, the market begins focusing more on expected loss (default probability times loss severity). So, bonds from two separate issuers with comparable (high) risk of default but different recovery rates may have similar ratings but trade at significantly different dollar prices.

Thus, bond investors who wait for rating agencies to change their ratings before making buy and sell decisions in their portfolios may be at risk of underperforming other investors who make portfolio decisions in advance of—or not solely based on—rating agency changes.

- **Rating agencies may make mistakes**. The mis-rating of billions of dollars of subprime-backed mortgage securities is one example. Other historical examples include the mis-ratings of US companies Enron and WorldCom and European issuer Parmalat. Like many investors, the rating agencies did not understand the accounting fraud being committed in those companies.
- **Some risks are difficult to capture in credit ratings**. Examples include litigation risk, such as that which can affect tobacco companies, or environmental and business risks faced by chemical companies and utility power plants. This

would also include the impact from natural disasters. Leveraged transactions, such as debt-financed acquisitions and large stock buybacks (share repurchases), are often difficult to anticipate and thus to capture in credit ratings.

As described, there are risks in relying on credit rating agency ratings when investing in bonds. Thus, while the credit rating agencies will almost certainly continue to play a significant role in the bond markets, it is important for investors to perform their own credit analyses and draw their own conclusions regarding the credit risk of a given debt issue or issuer.

EXAMPLE 4

Credit Ratings

1 Using the S&P rating scale, investment-grade bonds carry which of the following ratings?

   A AAA to EEE

   B BBB– to CCC

   C AAA to BBB–

2 Using both Moody's and S&P ratings, which of the following pairs of ratings is considered high yield, also known as "below investment grade," "speculative grade," or "junk"?

   A Baa1/BBB–

   B B3/CCC+

   C Baa3/BB+

3 What is the difference between an issuer rating and an issue rating?

   A The issuer rating applies to all of an issuer's bonds, whereas the issue rating considers a bond's seniority ranking.

   B The issuer rating is an assessment of an issuer's overall creditworthiness, whereas the issue rating is always higher than the issuer rating.

   C The issuer rating is an assessment of an issuer's overall creditworthiness, typically reflected as the senior unsecured rating, whereas the issue rating considers a bond's seniority ranking (e.g., secured or subordinated).

4 Based on the practice of notching by the rating agencies, a subordinated bond from a company with an issuer rating of BB would likely carry what rating?

   A B+

   B BB

   C BBB–

5 The fixed-income portfolio manager you work with asked you why a bond from an issuer you cover didn't rise in price when it was upgraded by Fitch from B+ to BB. Which of the following is the *most likely* explanation?

   A Bond prices never react to rating changes.

   B The bond doesn't trade often, so the price hasn't adjusted to the rating change yet.

   C The market was expecting the rating change, and so it was already "priced in" to the bond.

6 Amalgamated Corp. and Widget Corp. each have bonds outstanding with similar coupons and maturity dates. Both bonds are rated B2, B−, and B by Moody's, S&P, and Fitch, respectively. The bonds, however, trade at very different prices—the Amalgamated bond trades at €89, whereas the Widget bond trades at €62. What is the *most likely* explanation of the price (and yield) difference?

A Widget's credit ratings are lagging the market's assessment of the company's credit deterioration.

B The bonds have similar risks of default (as reflected in the ratings), but the market believes the Amalgamated bond has a higher expected loss in the event of default.

C The bonds have similar risks of default (as reflected in the ratings), but the market believes the Widget bond has a higher expected recovery rate in the event of default.

Solution to 1:

C is correct.

Solution to 2:

B is correct. Note that issuers with such ratings as Baa3/BB+ (answer C) are called "crossovers" because one rating is investment grade (the Moody's rating of Baa3) and the other is high yield (the S&P rating of BB+).

Solution to 3:

C is correct.

Solution to 4:

A is correct. The subordinated bond would have its rating notched lower than the company's BB rating, probably by two notches, reflecting the higher weight given to loss severity for below-investment-grade credits.

Solution to 5:

C is correct. The market was anticipating the rating upgrade and had already priced it in. Bond prices often do react to rating changes, particularly multi-notch ones. Even if bonds don't trade, their prices adjust based on dealer quotations given to bond pricing services.

Solution to 6:

A is correct. Widget's credit ratings are probably lagging behind the market's assessment of its deteriorating creditworthiness. Answers B and C both state the situation backwards. If the market believed that the Amalgamated bond had a higher expected loss given default, then that bond would be trading at a lower, not a higher, price. Similarly, if the market believed that the Widget bond had a higher expected recovery rate in the event of default, then that bond would be trading at a higher, not a lower, price.

TRADITIONAL CREDIT ANALYSIS: CORPORATE DEBT SECURITIES

5

f explain the four Cs (Capacity, Collateral, Covenants, and Character) of traditional credit analysis

The goal of credit analysis is to assess an issuer's ability to satisfy its debt obligations, including bonds and other indebtedness, such as bank loans. These debt obligations are contracts, the terms of which specify the interest rate to be paid, the frequency and timing of payments, the maturity date, and the covenants that describe the permissible and required actions of the borrower. Because corporate bonds are contracts, enforceable by law, credit analysts generally assume an issuer's willingness to pay and concentrate instead on assessing its ability to pay. Thus, the main focus in credit analysis is to understand a company's ability to generate cash flow over the term of its debt obligations. In so doing, analysts must assess both the credit quality of the company and the fundamentals of the industry in which the company operates. Traditional credit analysis considers the sources, predictability, and sustainability of cash generated by a company to service its debt obligations. This section will focus on corporate credit analysis; in particular, it will emphasize non-financial companies. Financial institutions have very different business models and funding profiles from industrial and utility companies.

5.1 Credit Analysis vs. Equity Analysis: Similarities and Differences

The description of credit analysis suggests credit and equity analyses should be very similar, and in many ways, they are. There are motivational differences, however, between equity and fixed-income investors that are an important aspect of credit analysis. Strictly speaking, management works for the shareholders of a company. Its primary objective is to maximize the value of the company for its owners. In contrast, management's legal duty to its creditors—including bondholders—is to meet the terms of the governing contracts. Growth in the value of a corporation from rising profits and cash flow accrues to the shareholders, while the best outcome for bondholders is to receive full, timely payment of interest and repayment of principal when due. Conversely, shareholders are more exposed to the decline in value if a company's earnings and cash flow decline because bondholders have a prior claim on cash flow and assets. But if a company's earnings and cash flow decline to the extent that it can no longer make its debt payments, then bondholders are at risk of loss as well.

In summary, in exchange for a prior claim on cash flow and assets, bondholders do not share in the growth in value of a company (except to the extent that its creditworthiness improves) but have downside risk in the event of default. In contrast, shareholders have theoretically unlimited upside opportunity, but in the event of default, their investment is typically wiped out before the bondholders suffer a loss. This is very similar to the type of payoff patterns seen in financial options. In fact, in recent years, credit risk models have been developed based on the insights of option pricing theory. Although it is beyond the scope of this present introduction to the subject, it is an expanding area of interest to both institutional investors and rating agencies.

Thus, although the analysis is similar in many respects for both equity and credit, equity analysts are interested in the strategies and investments that will increase a company's value and grow earnings per share. They then compare that earnings and growth potential with that of other companies in a given industry. Credit analysts will look more at the downside risk by measuring and assessing the sustainability of a company's cash flow relative to its debt levels and interest expense. Importantly for credit analysts, the balance sheet will show the composition of an issuer's debt—the overall amount, how much is coming due and when, and the distribution by seniority ranking. In general, equity analysts will focus more on income and cash flow statements, whereas credit analysts tend to focus more on the balance sheet and cash flow statements.

5.2 The Four Cs of Credit Analysis: A Useful Framework

Traditionally, many analysts evaluated creditworthiness based on what is often called the "four Cs of credit analysis":

- Capacity
- Collateral
- Covenants
- Character

Capacity refers to the ability of the borrower to make its debt payments on time; this is the focus of this section. **Collateral** refers to the quality and value of the assets supporting the issuer's indebtedness. **Covenants** are the terms and conditions of lending agreements that the issuer must comply with. **Character** refers to the quality of management. Each of these will now be covered in greater detail. Please note that the list of Cs is a convenient way to summarize the important aspects of the analysis; it is not a checklist to be applied mechanically, nor is it always exhaustive.

5.2.1 *Capacity*

Capacity is the ability of a borrower to service its debt. To determine that, credit analysis, in a process similar to equity analysis, starts with industry analysis and then turns to examination of the specific issuer (company analysis).

Industry structure. The Porter framework (Michael E. Porter, "The Five Competitive Forces That Shape Strategy," *Harvard Business Review* 86 [1; 2008]: 78–93) considers the effects of five competitive forces on an industry:

1 **Threat of entry.** Threat of entry depends on the extent of barriers to entry and the expected response from incumbents to new entrants. Industries with high entry barriers tend to be more profitable and have lower credit risk than industries with low entry barriers because incumbents do not need to hold down prices or take other steps to deter new entrants. High entry barriers can take many forms, including high capital investment, such as in aerospace; large, established distribution systems, such as in auto dealerships; patent protection, such as in technology or pharmaceutical industries; or a high degree of regulation, such as in utilities.
2 **Bargaining power of suppliers.** An industry that relies on just a few suppliers tends to be less profitable and to have greater credit risk than an industry that has multiple suppliers. Industries and companies with just a few suppliers have limited negotiating power to keep the suppliers from raising prices, whereas industries that have many suppliers can play them off against each other to keep prices in check.
3 **Bargaining power of customers.** Industries that rely heavily on just a few main customers have greater credit risk because the negotiating power lies with the buyers.

4 **Threat of substitutes.** Industries (and companies) that offer products and services that provide great value to their customers, and for which there are not good or cost-competitive substitutes, typically have strong pricing power, generate substantial cash flows, and represent less credit risk than other industries or companies. Certain (patent-protected) drugs are an example. Over time, however, disruptive technologies and inventions can increase substitution risk. For example, years ago, airplanes began displacing many trains and steamships. Newspapers were considered to have a nearly unassailable market position until television and then the internet became substitutes for how people received news and information. Over time, recorded music has shifted from records to tapes, to compact discs, to mp3s and other forms of digital media.

5 **Rivalry among existing competitors.** Industries with strong rivalry—because of numerous competitors, slow industry growth, or high barriers to exit—tend to have less cash flow predictability and, therefore, higher credit risk than industries with less competition. Regulation can affect the extent of rivalry and competition. For example, regulated utilities typically have a monopoly position in a given market, which results in relatively stable and predictable cash flows.

It is important to consider how companies in an industry generate revenues and earn profits. Is it an industry with high fixed costs and capital investment or one with modest fixed costs? These structures generate revenues and earn profits in very different ways. Two examples of industries with high fixed costs, also referred to as "having high operating leverage," are airlines and hotels. Many of their operating costs are fixed—running a hotel, flying a plane—so they cannot easily cut costs. If an insufficient number of people stay at a hotel or fly in a plane, fixed operating costs may not be covered and losses may result. With higher occupancy of a hotel or plane, revenues are higher, and it is more likely that fixed costs will be covered and profits earned.

Industry fundamentals. After understanding an industry's structure, the next step is to assess its fundamentals, including its sensitivity to macroeconomic factors, its growth prospects, its profitability, and its business need—or lack thereof—for high credit quality. Judgments about these can be made by looking at the following:

- *Cyclical or non-cyclical.* This is a crucial assessment because industries that are cyclical—that is, have greater sensitivity to broader economic performance—have more volatile revenues, margins, and cash flows and thus are inherently riskier than non-cyclical industries. Food producers, retailers, and health care companies are typically considered non-cyclical, whereas auto and steel companies can be very cyclical. Companies in cyclical industries should carry lower levels of debt relative to their ability to generate cash flow over an economic cycle than companies in less-cyclical or non-cyclical industries.
- *Growth prospects.* Although growth is typically a greater focus for equity analysts than for credit analysts, bond investors have an interest in growth as well. Industries that have little or no growth tend to consolidate via mergers and acquisitions. Depending upon how these are financed (e.g., using stock or debt) and the economic benefits (or lack thereof) of the merger, they may or may not be favorable to corporate bond investors. Weaker competitors in slow-growth industries may begin to struggle financially, adversely affecting their creditworthiness.
- *Published industry statistics.* Analysts can get an understanding of an industry's fundamentals and performance by researching statistics that are published by and available from a number of different sources, including the rating agencies, investment banks, industry publications, and frequently, government agencies.

Company fundamentals. Following analysis of an industry's structure and fundamentals, the next step is to assess the fundamentals of the company: the corporate borrower. Analysts should examine the following:

- Competitive position
- Track record/operating history
- Management's strategy and execution
- Ratios and ratio analysis
- When assessing the company fundamentals, analysts should also explore how the fundamentals are impacted by environmental, social, and governance (ESG) factors.

Competitive position. Based on their knowledge of the industry structure and fundamentals, analysts assess a company's competitive position within the industry. What is its market share? How has it changed over time: Is it increasing, decreasing, holding steady? Is it well above (or below) its peers? How does it compare with respect to cost structure? How might it change its competitive position? What sort of financing might that require?

Track record/Operating history. How has the company performed over time? It's useful to go back several years and analyze the company's financial performance, perhaps during times of both economic growth and contraction. What are the trends in revenues, profit margins, and cash flow? Capital expenditures represent what percentage of revenues? What are the trends on the balance sheet—use of debt versus equity? Was this track record developed under the current management team? If not, when did the current management team take over?

Management's strategy and execution. What is management's strategy for the company—to compete and to grow? Does it make sense, and is it plausible? How risky is it, and how differentiated is it from its industry peers? Is it venturing into unrelated businesses? Does the analyst have confidence in management's ability to execute? What is management's track record, both at this company and at previous ones?

Credit analysts also want to know and understand how management's strategy will affect its balance sheet. Does management plan to manage the balance sheet prudently, in a manner that doesn't adversely affect bondholders? Analysts can learn about management's strategy from reading comments, discussion, and analysis that are included with financial statements filed with appropriate regulators, listening to conference calls about earnings or other big announcements (e.g., acquisitions), going to company websites to find earnings releases and copies of slides of presentations at various industry conferences, visiting and speaking with the company, and so on.

EXAMPLE 5

Industry and Company Analysis

1 Given a hotel company, a chemical company, and a food retail company, which is *most likely* to be able to support a high debt load over an economic cycle?

A The hotel company, because people need a place to stay when they travel.

B The chemical company, because chemicals are a key input to many products.

   C The food retail company, because such products as food are typically resistant to recessions.

2 Heavily regulated monopoly companies, such as utilities, often carry high debt loads. Which of the following statements about such companies is *most* accurate?

   A Regulators require them to carry high debt loads.

   B They generate strong and stable cash flows, enabling them to support high levels of debt.

   C They are not very profitable and need to borrow heavily to maintain their plant and equipment.

3 XYZ Corp. manufactures a commodity product in a highly competitive industry in which no company has significant market share and where there are low barriers to entry. Which of the following *best* describes XYZ's ability to take on substantial debt?

   A Its ability is very limited because companies in industries with those characteristics generally cannot support high debt loads.

   B Its ability is high because companies in industries with those characteristics generally have high margins and cash flows that can support significant debt.

   C We don't have enough information to answer the question.

Solution to 1:

C is correct. Food retail companies are considered non-cyclical, whereas hotel and chemical companies are more cyclical and thus more vulnerable to economic downturns.

Solution to 2:

B is correct. Because such monopolies' financial returns are generally dictated by the regulators, they generate consistent cash flows and are thus able to support high debt levels.

Solution to 3:

A is correct. Companies in industries with those characteristics typically have low margins and limited cash flow and thus cannot support high debt levels.

---

g **calculate and interpret financial ratios used in credit analysis**

h **evaluate the credit quality of a corporate bond issuer and a bond of that issuer, given key financial ratios of the issuer and the industry**

Ratios and ratio analysis. To provide context to the analysis and understanding of a company's fundamentals—based on the industry in which it operates, its competitive position, its strategy and execution—a number of financial measures derived from the company's principal financial statements are examined. Credit analysts calculate a number of ratios to assess the financial health of a company, identify trends over time, and compare companies across an industry to get a sense of relative creditworthiness. Note that typical values of these ratios vary widely from one industry to another because of different industry characteristics previously identified: competitive structure, economic cyclicality, regulation, and so on.

We will categorize the key credit analysis measures into three different groups:

- Profitability and cash flow

- Leverage
- Coverage

Profitability and cash flow measures. It is from profitability and cash flow generation that companies can service their debt. Credit analysts typically look at operating profit margins and operating income to get a sense of a company's underlying profitability and see how it varies over time. Operating income is defined as operating revenues minus operating expenses and is commonly referred to as "earnings before interest and taxes" (EBIT). Credit analysts focus on EBIT because it is useful to determine a company's performance prior to costs arising from its capital structure (i.e., how much debt it carries versus equity). And "before taxes" is used because interest expense is paid before income taxes are calculated.

Several measures of cash flow are used in credit analysis; some are more conservative than others because they make certain adjustments for cash that gets used in managing and maintaining the business or in making payments to shareholders. The cash flow measures and leverage and coverage ratios discussed next are non-IFRS in the sense that they do not have official IFRS definitions; the concepts, names, and definitions given should be viewed as one usage among several possible, in most cases.

- **Earnings before interest, taxes, depreciation, and amortization (EBITDA).** EBITDA is a commonly used measure of cash flow that takes operating income and adds back depreciation and amortization expense because those are non-cash items. This is a somewhat crude measure of cash flow because it excludes certain cash-related expenses of running a business, such as capital expenditures and changes in (non-cash) working capital. Thus, despite its popularity as a cash flow measure, analysts look at other measures in addition to EBITDA.
- **Funds from operations (FFO).** Standard & Poor's defines funds from operations as net income from continuing operations plus depreciation, amortization, deferred income taxes, and other non-cash items.
- **Free cash flow before dividends (FCF before dividends).** This measures excess cash flow generated by the company (excluding non-recurring items) before payments to shareholders or that could be used to pay down debt or pay dividends. It can be calculated as net income (excluding non-recurring items) plus depreciation and amortization minus increase (plus decrease) in non-cash working capital minus capital expenditures. This is, depending upon the treatment of dividends and interest in the cash flow statement, approximated by the cash flow from operating activities minus capital expenditures. Companies that have negative free cash flow before payments to shareholders will be consuming cash they have or will need to rely on additional financing—from banks, bond investors, or equity investors. This obviously represents higher credit risk.
- **Free cash flow after dividends (FCF after dividends).** This measure just takes free cash flow before dividends and subtracts dividend payments. If this number is positive, it represents cash that could be used to pay down debt or build up cash on the balance sheet. Either action may be viewed as deleveraging, which is favorable from a credit risk standpoint. Some credit analysts will calculate net debt by subtracting balance sheet cash from total debt, although they shouldn't assume the cash will be used to pay down debt. Actual debt paid down from free cash flow is a better indicator of deleveraging. Some analysts will also deduct stock buybacks to get the "truest" measure of free cash flow that can be used to de-lever on either a gross or net debt basis; however, others view stock buybacks (share repurchases) as more discretionary and as having less certain timing than dividends, and thus treat those two types of shareholder payments differently when calculating free cash flow.

Leverage ratios. A few measures of leverage are used by credit analysts. The most common are the debt/capital, debt/EBITDA, and measures of funds or cash flows/debt ratios. Note that many analysts adjust a company's reported debt levels for debt-like liabilities, such as underfunded pensions and other retiree benefits, as well as operating leases. When adjusting for leases, analysts will typically add back the imputed interest or rent expense to various cash flow measures.

- **Debt/capital**. Capital is calculated as total debt plus shareholders equity. This ratio shows the percentage of a company's capital base that is financed with debt. A lower percentage of debt indicates lower credit risk. This traditional ratio is generally used for investment-grade corporate issuers. Where goodwill or other intangible assets are significant (and subject to obsolescence, depletion, or impairment), it is often informative to also compute the debt to capital ratio after assuming a write-down of the after-tax value of such assets.
- **Debt/EBITDA.** This ratio is a common leverage measure. Analysts use it on a "snapshot" basis, as well as to look at trends over time and at projections and to compare companies in a given industry. Rating agencies often use it as a trigger for rating actions, and banks reference it in loan covenants. A higher ratio indicates more leverage and thus higher credit risk. Note that this ratio can be very volatile for companies with high cash flow variability, such as those in cyclical industries and with high operating leverage (fixed costs).
- **FFO/debt.** Credit rating agencies often use this leverage ratio. They publish key median and average ratios, such as this one, by rating category so analysts can get a sense of why an issuer is assigned a certain credit rating, as well as where that rating may migrate based on changes to such key ratios as this one. A higher ratio indicates greater ability to pay debt by funds from operations.
- **FCF after dividends/debt**. A higher ratio indicates that a greater amount of debt can be paid off from free cash flow after dividend payments.

Coverage ratios. Coverage ratios measure an issuer's ability to meet—to "cover"—its interest payments. The two most common are the EBITDA/interest expense and EBIT/interest expense ratios.

- **EBITDA/interest expense.** This measurement of interest coverage is a bit more liberal than the one that uses EBIT because it does not subtract out the impact of (non-cash) depreciation and amortization expense. A higher ratio indicates higher credit quality.
- **EBIT/interest expense.** Because EBIT does not include depreciation and amortization, it is considered a more conservative measure of interest coverage. This ratio is now used less frequently than EBITDA/interest expense.

Exhibit 9 is an example of key average credit ratios by rating category for industrial companies over the 12-month period 3Q2017–3Q2018, as calculated by Bloomberg Barclays Indices, using public company data. Note only a few AAA-rated corporations remain, so the small sample size can skew the average ratios of a few key credit metrics. That said, it should be clear that, overall, higher-rated issuers have stronger credit metrics.

Exhibit 9 Industrial Comparative Ratio Analysis

Credit Rating	EBITDA Margin (%)	Return on Capital (%)	EBIT Interest Coverage (x)	EBITDA Interest Coverage (x)	FFO/Debt (%)	Free Operations Cash Flow/Debt (%)	Debt/ EBITDA (x)	Debt/Debt plus Equity (%)
Aaa								
US	66.4	6.5	4.2	21.3	51.9	43.5	–0.2	43.3
Aa								
US	21.9	10.8	15.4	45.0	109.9	58.1	1.2	50.6
A								
US	26.0	13.5	13.3	18.9	49.1	31.8	1.8	51.2
Baa								
US	23.9	11.5	7.2	NA	40.7	20.3	3.9	49.4
Ba								
US	21.7	3.5	4.6	NA	27.7	11.0	4.1	64.0
B								
US	21.2	3.7	2.5	NA	20.3	1.8	5.2	69.3
Caa								
US	16.0	0.2	–0.6	1.3	10.0	–6.6	9.3	95.3

Note: As of 19 December 2018.
Source: Bloomberg Barclays Indices.

Comments on issuer liquidity. An issuer's access to liquidity is also an important consideration in credit analysis. Companies with high liquidity represent lower credit risk than those with weak liquidity, other factors being equal. The global financial crisis of 2008–2009 showed that access to liquidity via the debt and equity markets should not be taken for granted, particularly for companies that do not have strong balance sheets or steady operating cash flow.

When assessing an issuer's liquidity, credit analysts tend to look at the following:

- **Cash on the balance sheet.** Cash holdings provide the greatest assurance of having sufficient liquidity to make promised payments.
- **Net working capital.** The big US automakers used to have enormous negative working capital, despite having high levels of cash on the balance sheet. This proved disastrous when the global financial crisis hit in 2008 and the economy contracted sharply. Auto sales—and thus revenues—fell, the auto companies cut production, and working capital consumed billions of dollars in cash as accounts payable came due when the companies most needed liquidity.
- **Operating cash flow.** Analysts will project this figure out a few years and consider the risk that it may be lower than expected.
- **Committed bank lines.** Committed but untapped lines of credit provide contingent liquidity in the event that the company is unable to tap other, potentially cheaper, financing in the public debt markets.
- **Debt coming due and committed capital expenditures in the next one to two years.** Analysts will compare the sources of liquidity with the amount of debt coming due as well as with committed capital expenditures to ensure that companies can repay their debt and still invest in the business if the capital markets are somehow not available.

As will be discussed in more detail in the section on special considerations for high-yield credits, issuer liquidity is a bigger consideration for high-yield companies than for investment-grade companies.

EXAMPLE 6

Mallinckrodt PLC (Mallinckrodt) is an Ireland-incorporated specialty pharmaceutical company. As a credit analyst, you have been asked to assess its creditworthiness—on its own, compared to a competitor in its overall industry, and compared with a similarly rated company in a different industry. Using the financial statements provided in Exhibits 10 through 12 for the three years ending 31 December 2015, 2016, and 2017, address the following:

1 Calculate Mallinckrodt's operating profit margin, EBITDA, and free cash flow after dividends. Comment on what these measures indicate about Mallinckrodt's profitability and cash flow.

2 Determine Mallinckrodt's leverage ratios: debt/EBITDA, debt/capital, free cash flow after dividends/debt. Comment on what these leverage ratios indicate about Mallinckrodt's creditworthiness.

3 Calculate Mallinckrodt's interest coverage using both EBIT and EBITDA. Comment on what these coverage ratios indicate about Mallinckrodt's creditworthiness.

4 Using the credit ratios provided in Exhibit 11 on Johnson & Johnson, compare the creditworthiness of Mallinckrodt relative to Johnson & Johnson.

5 Compare the Exhibit 12 credit ratios of Luxembourg-based ArcelorMittal, one of the world's largest global steelmakers, with those of Mallinckrodt. Comment on the volatility of the credit ratios of the two companies. Which company looks to be more cyclical? What industry factors might explain some of the differences? In comparing the creditworthiness of these two companies, what other factors might be considered to offset greater volatility of credit ratios?

Exhibit 10a Mallinckrodt PLC Financial Statements

Consolidated Statements of Operations

(Dollars in millions, except per share amounts)	Year End 30 Sept.* 2015	Years Ended 31 Dec. 2016	2017
Net revenues	**2,923.1**	**3,399.5**	**3,221.6**
Operating expenses:			
Cost of sales	1,300.2	1,549.6	1,565.3
Research and development	203.3	267.0	277.3
Selling, general, and administrative expenses	1,023.8	1,070.3	920.9
Restructuring charges, net	45.0	33.0	31.2
Non-restructuring impairment charges	—	231.2	63.7
Gain on divestiture and license	(3.0)	—	(56.9)
Total operating expenses	**2,569.3**	**3,151.1**	**2,801.5**

(continued)

Exhibit 10a (Continued)

Consolidated Statements of Operations

(Dollars in millions, except per share amounts)	Year End 30 Sept.* 2015	Years Ended 31 Dec. 2016	2017
Operating income	**353.8**	**248.4**	**420.1**
Other (expense) income:			
Interest expense	(255.6)	(378.1)	(369.1)
Interest income	1.0	1.6	4.6
Other income (expense), net	8.1	(3.5)	6.0
Total other (expense) income, net	**(246.5)**	**(380.0)**	**(358.5)**
Income before income taxes and non-controlling interest	**107.3**	**(131.6)**	**61.6**
Provision (Benefit) for income taxes	(129.3)	(340.0)	(1,709.6)
Net income	**236.6**	**208.4**	**1,771.2**
Income from discontinued operations, net of income taxes	88.1	71.0	363.2
Net income attributable to common shareholders	**324.7**	**279.4**	**2,134.4**

*Mallinckrodt changed their fiscal year end from 30 September to 31 December.
Source: Company filings, Loomis, Sayles & Company.

Exhibit 10b Mallinckrodt PLC Financial Statements

Consolidated Balance Sheets

(Dollars in millions)	Year End 30 Sept.* 2015	Years Ended 31 Dec. 2016	2017
ASSETS			
Current assets:			
Cash and cash equivalents	365.9	342.0	1,260.9
Accounts receivable	489.6	431.0	445.8
Inventories	262.1	350.7	340.4
Deferred income taxes	139.2	—	—
Prepaid expenses and other current assets	194.4	131.9	84.1
Notes receivable	—	—	154.0
Current assets held for sale	394.9	310.9	—

Exhibit 10b (Continued)

Consolidated Balance Sheets

(Dollars in millions)	Year End 30 Sept.* 2015	Years Ended 31 Dec. 2016	2017
Total current assets	**1,846.1**	**1,566.5**	**2,285.2**
Property, plant, and equipment, net	793.0	881.5	966.8
Goodwill	3,649.4	3,498.1	3,482.7
Intangible assets, net	9,666.3	9,000.5	8,375.0
Other assets	225.7	259.7	171.2
Long-term assets held for sale	223.6	—	—
Total assets	**16,404.1**	**15,206.3**	**15,280.9**
LIABILITIES AND EQUITY			
Current liabilities:			
Current maturities of long-term debt	22.0	271.2	313.7
Accounts payable	116.8	112.1	113.3
Accrued payroll and payroll-related costs	95.0	76.1	98.5
Accrued interest	80.2	68.7	57.0
Income taxes payable	—	101.7	15.8
Accrued and other current liabilities	486.1	557.1	452.1
Current liabilities held for sale	129.3	120.3	—
Total current liabilities	**929.4**	**1,307.2**	**1,050.4**
Long-term debt	6,474.3	5,880.8	6,420.9
Pension and post-retirement benefits	114.2	136.4	67.1
Environmental liabilities	73.3	73.0	73.2
Deferred income taxes	3,117.5	2,398.1	689.0
Other income tax liabilities	121.3	70.4	94.1
Other liabilities	209.0	356.1	364.2
Long-term liabilities held for sale	53.9	—	—
Total liabilities	**11,092.9**	**10,222.0**	**8,758.9**
Shareholders' equity:			
Ordinary shares	23.5	23.6	18.4
Ordinary shares held in treasury at cost	(109.7)	(919.8)	(1,564.7)
Additional paid-in capital	5,357.6	5,424.0	5,492.6
Retained earnings	38.9	529.0	2,588.6
Accumulated other comprehensive income	0.9	(72.5)	(12.9)

(continued)

Exhibit 10b (Continued)

Consolidated Balance Sheets

(Dollars in millions)	Year End 30 Sept.* 2015	Years Ended 31 Dec. 2016	2017
Total shareholders' equity	5,311.2	4,984.3	6,522.0
Total liabilities and shareholders' equity	16,404.1	15,206.3	15,280.9

*Mallinckrodt changed their fiscal year end from 30 September to 31 December.
Source: Company filings, Loomis, Sayles & Company

Exhibit 10c Mallinckrodt PLC Financial Statements

Consolidated Statements of Cash Flow

(Dollars in millions)	Year End 30 Sept.* 2015	Years Ended 31 Dec. 2016	2017
Cash Flows from Operating Activities:			
Net income (loss)	324.7	279.4	2,134.4
Depreciation and amortization	672.5	831.7	808.3
Share-based compensation	117.0	45.4	59.2
Deferred income taxes	(191.6)	(528.3)	–1744.1
Non-cash impairment charges	—	231.2	63.7
Inventory provisions	—	8.5	34.1
Gain on disposal of discontinued operations	—	1.7	–418.1
Other non-cash items	(25.5)	45.5	–21.4
Change in working capital	33.4	153.7	–188.8
Net cash from operating activities	**930.5**	**1,068.8**	**727.3**
Cash Flows from Investing Activities:			
Capital expenditures	(148.0)	(199.1)	–186.1
Acquisitions and intangibles, net of cash acquired	(2,154.7)	(247.2)	–76.3
Proceeds from divestitures, net of cash	—	3.0	576.9
Other	3.0	(4.9)	3.9
Net cash from investing activities	**(2,299.7)**	**(448.2)**	**318.4**
Cash Flows from Financing Activities:			
Issuance of external debt	3,010.0	226.3	1,465
Repayment of external debt and capital leases	(1,848.4)	(525.7)	–917.2
Debt financing costs	(39.9)	—	–12.7
Proceeds from exercise of share options	34.4	10.8	4.1

Exhibit 10c (Continued)

Consolidated Statements of Cash Flow

	Year End 30 Sept.*	Years Ended 31 Dec.	
(Dollars in millions)	**2015**	**2016**	**2017**
Repurchase of shares	(92.2)	(536.3)	–651.7
Other	(28.1)	(21.8)	–17.7
Net cash from financing activities	**1,035.8**	**(846.7)**	**–130.2**
Effect of currency rate changes on cash	(11.6)	(1.2)	**2.5**
Net increase (decrease) in cash and cash equivalents	(345.0)	(227.3)	918
Cash and cash equivalents at beginning of period	777.6	588.4	361.1
Cash and cash equivalents at end of period	432.6	361.1	1,279.1

*Mallinckrodt changed their fiscal year end from 30 September to 31 December.
Source: Company filings, Loomis, Sayles & Company.

Exhibit 10d Mallinckrodt PLC Credit Ratios

	2015	2016	2017
Operating margin	12.1%	7.3%	13.0%
Debt/EBITDA	6.3x	5.7x	5.5x
EBITDA/Interest	4.0x	2.9x	3.3x
FCF/Debt	12.0%	14.1%	8.0%
Debt/Capital	55.0%	55.2%	50.8%

Source: Company filings, Loomis, Sayles & Company.

Exhibit 11 Johnson & Johnson's Credit Ratios

	2015	2016	2017
Operating profit margin	26.2%	29.5%	25.8%
Debt/EBITDA	0.9x	1.1x	1.4x
EBITDA/Interest	40.1x	34.4x	27.1x
FCF after dividends/Debt	81.1%	57.3%	51.4%
Debt/Capital	22.0%	28.1%	36.5%

Source: Company filings, Loomis, Sayles & Company.

Exhibit 12 ArcelorMittal Credit Ratios

	2015	2016	2017
Operating profit margin	0.3%	5.5%	7.7%
Debt/EBITDA	5.8x	2.3x	1.6x
EBITDA/Interest	2.5x	5.0x	9.2x
FCF after dividends/Debt	−2.8%	1.9%	13.5%
Debt/Capital	41.8%	29.7%	24.0%

Source: Company filings, Loomis, Sayles & Company.

Solutions:

1 Operating profit margin (%) = Operating income/Revenue

2015: 353.8/2,923.1 = 0.121 or 12.1%

2016: 248.4/3,399.5 = 0.073 or 7.3%

2017: 420.1/3,221.6 = 0.130 or 13.0%

EBITDA = Operating income + Depreciation and Amortization

2015: 353.8 + 672.5 = 1,026.3

2016: 248.4+ 831.7 = 1,080.1

2017: 420.1 + 808.3= 1,228.4

FCF after dividends = Cash flow from operations − Capital expenditures − Dividends

2015: 930.5 − 148 − 0 = 782.5

2016: 1,068.8 − 199.1 − 0 = 869.7

2017: 727.3 − 186.1 − 0 = 541.2

Operating profit margin decreased from 2015 to 2016 but increased from 2016 to 2017. Conversely, FCF after dividends increased from 2015 to 2016 but decreased from 2016 to 2017. EBITDA increased from 2015 to 2017. From 2015 to 2016, sales increased by 16.3% and operating expenses increased by 22.6%. As a result, operating profit margin decreased even though EBITDA and FCF increased. However, from 2016 to 2017, sales decreased by 5.2% and operating expenses decreased by 11.1%. As a result, operating profit margin and EBITDA increased, while FCF after dividends decreased.

2 Debt/EBITDA

Total debt = Short-term debt and Current portion of long-term debt + Long-term debt

2015: Debt: 22.0 + 6,474.3 = 6,496.3

Debt/EBITDA: 6,496.3/1,026.3 = 6.3x

2016: Debt: 271.2 + 5,880.8 = 6,152.0

Debt/EBITDA: 6,152.0/1,080.1 = 5.7x

2017: Debt: 313.7 + 6,420.9 = 6,734.6

Debt/EBITDA: 6,734.6/1,228. 4 = 5.5x

Debt/Capital (%)

Capital = Debt + Equity

2015: Capital: 6,496.3 + 5,311.2 = 11,807.5

Debt/Capital: 6,496.3/11,807.5 = 55.0%

2016: Capital: 6,152.0 + 4,984.3 = 11,136.3

Debt/Capital: 6,152.0/11,136.3 = 55.2%

2017: Capital: 6,734.6 + 6,522.0 = 13,256.6

Debt/Capital: 6,734.6/13,256.6 = 50.8%

FCF after dividends/Debt (%)

2015: 782.5/6,496.3 = 12.0%

2016: 869.7/6,152 = 14.1%

2017: 541.2/6,734.6 = 8.0%

Although debt/EBITDA and debt/capital improved between 2015 and 2017, the "FCF after dividends/Debt" deteriorated significantly as cash flow from operations declined as a result of the loss taken on disposal of discontinued operations. Given that the loss is most likely a non-recurring event, Mallinckrodt's creditworthiness likely improved over the 2015 to 2017 period.

3 EBIT/Interest expense

2015: 353.8/255.6 = 1.4x

2016: 248.4/378.1 = 0.7x

2017: 420.1/369.1 = 1.1x

EBITDA/Interest expense

2015: 1,026.3/255.6 = 4.0x

2016: 1,080.1/378.1= 2.9x

2017: 1,228.4/369.1 = 3.3x

Based on these coverage ratios, Mallinckrodt's creditworthiness declined from 2015 to 2016 and then showed modest improvement in 2017. The 2017 coverage ratios are still weaker than the 2015 coverage ratios, indicating that growth in EBIT and EBITDA are not keeping pace with the rising interest expense.

4 Johnson & Johnson (J&J) has a higher operating profit margin, better leverage ratios—lower debt/EBITDA, higher FCF after dividends/debt over the three years, lower debt/capital, and better interest coverage as measured by EBITDA/interest. Collectively, those ratios suggest J&J has higher credit quality than Mallinckrodt.

5 Mallinckrodt has both a higher and a less volatile operating profit margin than ArcelorMittal (Arcelor). However, while Mallinckrodt's leverage ratios have been deteriorating, Arcelor's have been improving. Based on the volatility of its cash flow and operating profit margin, Arcelor appears to be a much more cyclical credit. However, with its meaningfully lower debt levels, one could expect Arcelor to have a higher credit rating.

A steelmaker likely has a significant amount of long-term assets financed by debt. It is a highly competitive industry with little ability to distinguish products from other competitors. To mitigate the impact of its more volatile credit ratios, Arcelor might maintain high levels of liquidity. Its size and global diversity may also be a "plus." Given its size, it may be able to negotiate favorable supplier and customer contracts and keep costs down through economies of scale.

5.2.2 *Collateral*

Collateral, or asset value, analysis is typically emphasized more with lower credit quality companies. As discussed earlier, credit analysts focus primarily on probability of default, which is mostly about an issuer's ability to generate sufficient cash flow to support its debt payments, as well as its ability to refinance maturing debt. Only when the default probability rises to a sufficient level do analysts typically consider asset or collateral value in the context of loss severity in the event of default.

Analysts do think about the value and quality of a company's assets; however, these are difficult to observe directly. Factors to consider include the nature and amount of intangible assets on the balance sheet. Some assets, such as patents, are clearly valuable and can be sold if necessary to cover liabilities. Goodwill, on the other hand, is not considered a high-quality asset. In fact, sustained weak financial performance most likely implies that a company's goodwill will be written down, reinforcing its poor quality. Another factor to consider is the amount of depreciation an issuer takes relative to its capital expenditures: Low capital expenditures relative to depreciation expense could imply that management is insufficiently investing in its business, which will lead to lower-quality assets, potentially reduced future operating cash flow, and higher loss severity in the event of default.

A market-based signal that credit analysts use to impute the quality of a publicly traded company's assets, and its ability to support its debt, is equity market capitalization. For instance, a company whose stock trades below book value may have lower-quality assets than is suggested by the amount reported on the balance sheet.

As economies become more service- and knowledge-based and those types of companies issue debt, it's important to understand that these issuers rely more on human and intellectual capital than on "hard assets." In generating profits and cash flow, these companies are not as asset intensive. One example would be software companies. Another example would be investment management firms. Human- and intellectual-capital-based companies may generate a lot of cash flow, but their collateral value is questionable unless there are patents and other types of intellectual property and "intangible capital" that may not appear directly on the balance sheet but could be valuable in the event of financial distress or default.

Regardless of the nature of the business, the key point of collateral analysis is to assess the value of the assets relative to the issuer's level—and seniority ranking—of debt.

5.2.3 *Covenants*

Covenants are meant to protect creditors while also giving management sufficient flexibility to operate its business on behalf of and for the benefit of the shareholders. They are integral to credit agreements, whether they are bonds or bank loans, and they spell out what the issuer's management is (1) obligated to do and (2) limited in doing. The former are called "affirmative covenants," whereas the latter are called "negative" or "restrictive covenants." Obligations would include such duties as making interest and principal payments and filing audited financial statements on a timely basis. Covenants might also require a company to redeem debt in the event of the company being acquired, the change of control covenant, or to keep the ratio of debt to EBITDA below some prescribed amount. The limitations might include a cap on the amount of cash that can be paid out to shareholders relative to earnings, or perhaps on the amount of additional secured debt that can be issued. Covenant violations are a breach of contract and can be considered default events unless they are cured in a short time or a waiver is granted.

For corporate bonds, covenants are described in the bond **prospectus**, the document that is part of a new bond issue. The prospectus describes the terms of the bond issue, as well as supporting financial statements, to help investors perform their analyses and make investment decisions as to whether or not to submit orders to buy the new bonds. Actually, the **trust deed** or **bond indenture** is the governing legal credit agreement and is typically incorporated by reference in the prospectus.

Covenants are an important but underappreciated part of credit analysis. Strong covenants protect bond investors from the possibility of management taking actions that would hurt an issuer's creditworthiness. For example, without appropriate covenants, management might pay large dividends, undertake stock buybacks well in excess of free cash flow, sell the company in a leveraged buyout, or take on a lot of secured debt that structurally subordinates unsecured bondholders. All of these actions would enrich shareholders at the expense of bondholders. Recall that management works for the shareholders and that bonds are contracts, with management's only real obligation to creditors being to uphold the terms of the contract. The inclusion of covenants in the contract is intended to protect bondholders.

The bond-buying investor base is very large and diverse, particularly for investment-grade debt. It includes institutional investors, such as insurance companies, investment management firms, pension funds, mutual funds, hedge funds, sovereign wealth funds, and so on. Although there are some very large institutional investors, the buyer base is fragmented and does not—and legally cannot—act as a syndicate. Thus, bondholders are generally not able to negotiate strong covenants on most new bond issues. At the same time, issuers expect that borrowing from capital markets will carry lighter convenants than borrowing from banks. During weak economic or market conditions, however, covenants on new bond issues tend to be stronger because investors require additional incentive to lend money while seeking more protection. A few organized institutional investor groups are focused on strengthening covenants: the Credit Roundtable (thecreditroundtable.org) in the United States and the European Model Covenant Initiative in the United Kingdom.

Covenant language is often very technical and written in "legalese," so it can be helpful to have an in-house person with a legal background to review and interpret the specific covenant terms and wording. One might also use a third-party service specializing in covenant analysis, such as Covenant Review (www.covenantreview.com).

We will go into more detail on specific covenants in the section on special considerations for high-yield bonds.

5.2.4 *Character*

The character of a corporate borrower can be difficult to observe. The analysis of character as a factor in credit analysis dates to when loans were made to companies owned by individuals. Most corporate bond issuers are now publicly owned by shareholders or privately owned by pools of capital, such as private equity firms. Management often has little ownership in a corporation, so analysis and assessment of character is different than it would be for owner-managed firms. Credit analysts can make judgments about management's character in the following ways:

- An assessment of the soundness of management's strategy.
- Management's track record in executing past strategies, particularly if they led to bankruptcy or restructuring. A company run by executives whose prior positions/ventures resulted in significant distress might still be able to borrow in the debt markets, but it would likely have to borrow on a secured basis and/or pay a higher interest rate.
- Use of aggressive accounting policies and/or tax strategies. Examples might include using a significant amount of off-balance-sheet financing, capitalizing versus immediately expensing items, recognizing revenue prematurely, and/or frequently changing auditors. These are potential warning flags to other behaviors or actions that may adversely impact an issuer's creditworthiness.
- Any history of fraud or malfeasance—a major warning flag to credit analysts.

- Previous poor treatment of bondholders—for example, management actions that resulted in major credit rating downgrades. These actions might include a debt-financed acquisition, a large special dividend to shareholders, or a major debt-financed stock buyback program.

EXAMPLE 7

The Four Cs

1 Which of the following would not be a bond covenant?

   A The issuer must file financial statements with the bond trustee on a timely basis.

   B The company can buy back as much stock as it likes.

   C If the company offers security to any creditors, it must offer security to this bond issue.

2 Why should credit analysts be concerned if a company's stock trades below book value?

   A It means the company is probably going bankrupt.

   B It means the company will probably incur lots of debt to buy back its undervalued stock.

   C It's a signal that the company's asset value on its balance sheet may be impaired and have to be written down, suggesting less collateral protection for creditors.

3 If management is of questionable character, how can investors incorporate this assessment into their credit analysis and investment decisions?

   A They can choose not to invest based on the increased credit risk.

   B They can insist on getting collateral (security) and/or demand a higher return.

   C They can choose not to invest or insist on additional security and/or higher return.

Solution to 1:

B is correct. Covenants describe what the borrower is (1) obligated to do or (2) limited in doing. It's the absence of covenants that would permit a company to buy back as much stock as it likes. A requirement that the company offer security to this bond issue if it offers security to other creditors (answer C) is referred to as a "negative pledge."

Solution to 2:

C is correct.

Solution to 3:

C is correct. Investors can always say no if they are not comfortable with the credit risk presented by a bond or issuer. They may also decide to lend to a borrower with questionable character only on a secured basis and/or demand a higher return for the perceived higher risk.

CREDIT RISK VS. RETURN: YIELDS AND SPREADS

6

i. describe macroeconomic, market, and issuer-specific factors that influence the level and volatility of yield spreads

The material in this section applies to all bonds subject to credit risk. For simplicity, in what follows all such bonds are sometimes referred to as "corporate" bonds.

As in other types of investing, taking more risk in credit offers higher potential return but with more volatility and less certainty of earning that return. Using credit ratings as a proxy for risk, Exhibit 13 shows the composite yield-to-maturity for bonds of all maturities within each rating category in the US and European bond markets according to Bloomberg Barclays, one of the largest providers of fixed-income market indexes. For non-investment-grade bonds, the yield-to-call (YTC) or yield-to-worst (YTW) is reported as many contain optionality.

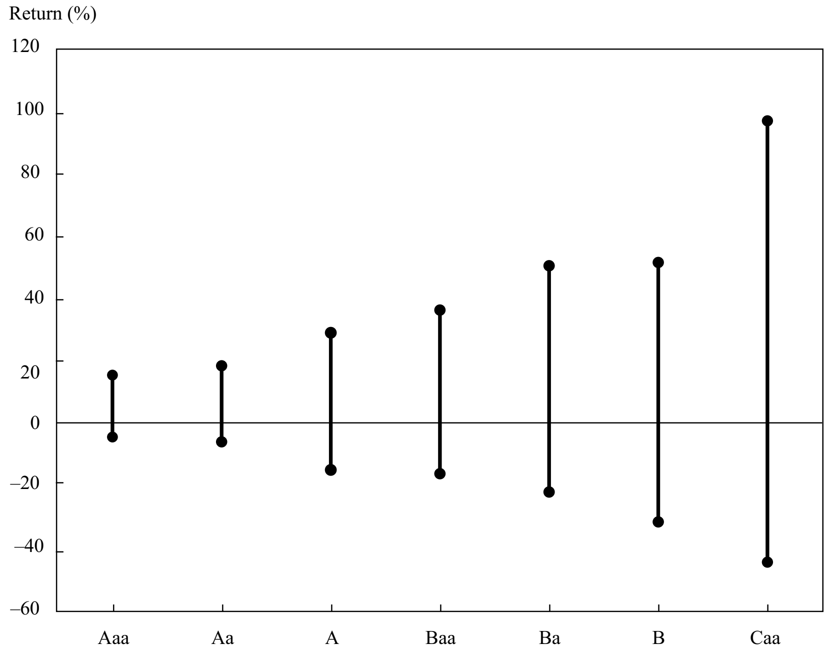
Exhibit 13 Corporate Yields by Rating Category (%)

Bloomberg Barclays Indices	Investment Grade				Non-Investment Grade			
	AAA	AA	A	BBB	BB	B	CCC	CC–D
US	3.63	3.52	3.86	4.35	5.14	6.23	8.87	19.51
Pan European*	1.25	0.76	1.18	1.67	2.92	5.63	8.78	54.95

*Pan European yields may be "artificially low" due to the ECB's extraordinary corporate bond Quantitative Easing (QE) Program.
Note: Data as of 30 September 2018.
Source: Bloomberg Barclays Indices.

Note that the lower the credit quality, the higher the quoted yield. The realized yield, or return, will almost always be different because of changes in interest rates, re-investment of coupons, holding period changes, and the credit-related risks discussed earlier. In addition to the absolute returns, the volatility of returns will also vary by rating. Trailing 12-month returns by credit rating category and the volatility (standard deviation) of those returns are shown in Exhibit 14.

Exhibit 14 US Credit Trailing 12-Month Returns by Rating Category, 31 December 1996–30 September 2018



	Aaa	Aa	A	Baa	Ba	B	Caa
Avg. Return:	4.82%	5.60%	5.78%	6.53%	7.98%	6.79%	8.26%
Max. Return:	16.08%	19.30%	29.39%	36.43%	50.80%	52.10%	97.47%
Min. Return:	–4.17%	–5.47%	–14.98%	–15.70%	–22.12%	–31.47%	–44.35%
Avg. Volatility:	3.43%	4.14%	4.76%	5.15%	5.75%	7.13%	11.00%

Sources: Bloomberg Barclays Indices and Loomis, Sayles & Company.

As shown in the exhibit, the higher the credit risk, the greater the return potential and the higher the volatility of that return. This pattern is consistent with other types of investing that involve risk and return (although average returns on single-B rated bonds appear anomalous in this example).

For extremely liquid bonds that are deemed to have virtually no default risk (e.g., German government bonds, or *Bunds*), the yield for a given maturity is a function of real interest rates plus an expected inflation rate premium. The yield on corporate bonds will include an additional risk premium that provides the investor with compensation for the credit and liquidity risks and possibly the tax impact of holding a specific bond. Changes in any of these components will alter the yield, price, and return on the bond. In general, however, it is not possible to directly observe the market's assessment of the components separately—analysts can only observe the total yield spread.

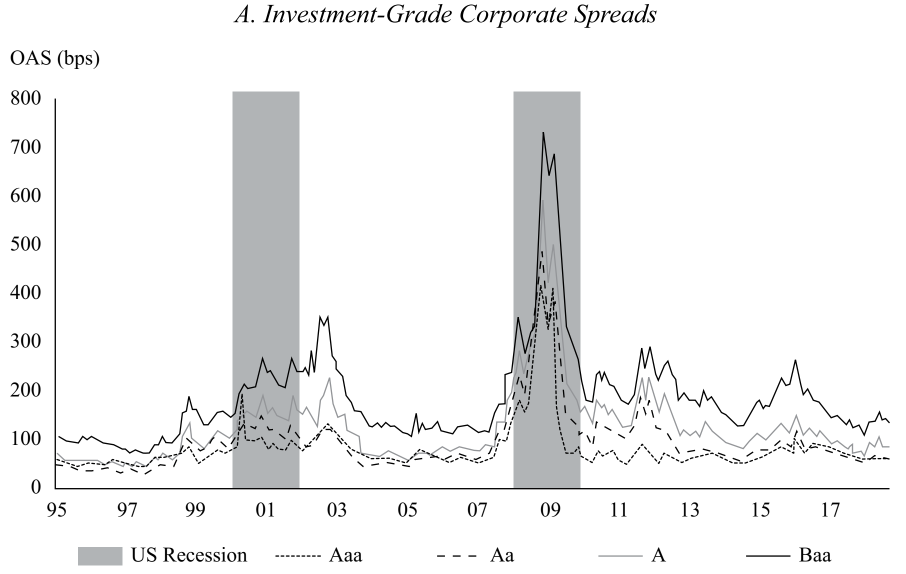
Spreads on all corporate bonds can be affected by a number of factors, with lower-quality issuers typically experiencing greater spread volatility. These factors, which are frequently linked, include the following:

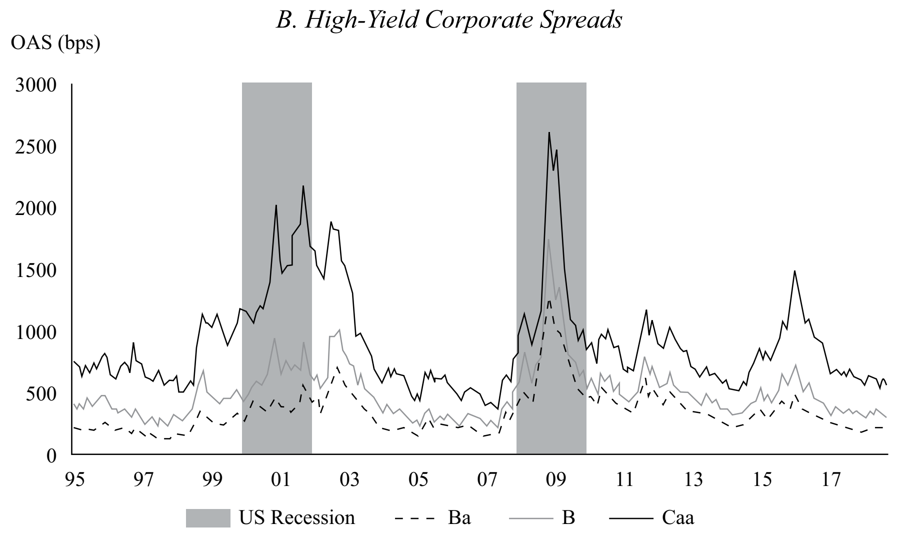
- **Credit cycle.** As the credit cycle improves, credit spreads will narrow. Conversely, a deteriorating credit cycle will cause credit spreads to widen. Spreads are tightest at or near the top of the credit cycle, when financial markets believe risk is low; they are widest at or near the bottom of the credit cycle, when financial markets believe risk is high.

- **Broader economic conditions.** Not surprisingly, weakening economic conditions will push investors to desire a greater risk premium and drive overall credit spreads wider. Conversely, a strengthening economy will cause credit spreads to narrow because investors anticipate credit measures will improve due to rising corporate cash flow, thus reducing the risk of default. In a steady, low-volatility environment, credit spreads will typically also narrow as investors tend to "reach for yield."
- **Funding availability in financial sector.** Bonds trade primarily over the counter, so investors need broker–dealers to commit capital for market-making purposes. Episodes of financial and regulatory stress have the potential to greatly reduce the total capital available for making markets (to facilitate trading) and the willingness to buy/sell credit-risky bonds. Future regulatory reform may well lead to persistent or even permanent reductions in broker-provided capital. Funding stresses would naturally translate into wider spreads.
- **General market supply and demand.** In periods of heavy new issue supply, credit spreads will widen if there is insufficient demand. In periods of high demand for bonds, spreads will move tighter.
- **Financial performance of the issuer.** Corporate bond spreads will be impacted by earnings releases, news, and other developments associated with the issuer. Earlier we explained the "four Cs of credit analysis": capacity, collateral, covenants, and character. Announcement and disclosures by the issuer will impact investors' view of the issuer's financial performance and ability to service and repay its debt. Good news increases the attractiveness of buying and holding bonds issued by the corporation, which raises bond prices and narrows spreads; bad news will have the opposite effect.

A number of these factors played a role during the global financial crisis of 2008–2009, causing spreads to widen dramatically, as shown in Exhibit 15, before narrowing sharply as governments intervened and markets stabilized. This is shown in two panels—one for investment grade, another for high yield—because of the much greater spread volatility in high-yield bonds, particularly CCC rated credits. This spread volatility is reflected in the different spread ranges on the *y*-axes. OAS is option-adjusted spread, which incorporates the value of the embedded call option in certain corporate bonds that issuers have the right to exercise before maturity.

Exhibit 15 US Investment-Grade and High-Yield Corporate Spreads





Sources: Bloomberg Barclays Indices and Loomis, Sayles & Company.

EXAMPLE 8

Yield Spreads

1 Which bonds are likely to exhibit the greatest spread volatility?

A Bonds from issuers rated AA

B Bonds from issuers rated BB

C Bonds from issuers rated A

2 If investors become increasingly worried about the economy—say, as shown by declining stock prices—what is the *most likely* impact on credit spreads?

A There will be no change to credit spreads. They aren't affected by equity markets.

B Narrower spreads will occur. Investors will move out of equities into debt securities.

C Wider spreads will occur. Investors are concerned about weaker creditworthiness.

Solution to 1:

B is correct. Lower-quality bonds exhibit greater spread volatility than higher-quality bonds. All of the factors that affect spreads—the credit cycle, economic conditions, financial performance, market-making capacity, and supply/demand conditions—will tend to have a greater impact on the pricing of lower-quality credits.

Solution to 2:

C is correct. Investors will require higher yields as compensation for the credit deterioration—including losses—that is likely to occur in a weakening economy.

6.1 Credit Risk vs. Return: The Price Impact of Spread Changes

We have discussed how yield spreads on credit-risky debt obligations, such as corporate bonds, can fluctuate based on a number of factors, including changes in the market's view of issuer-specific or idiosyncratic risk. The next question to consider is how these spread changes affect the price of and return on these bonds.

Although bond investors do concern themselves with default risks, recall that the probability of default for higher-quality bonds is typically very low: For investment-grade bonds, annual defaults are nearly always well below 1% (recall Exhibit 6). On the other hand, default rates can be very high for lower-quality issuers, although they can vary widely depending upon the credit cycle, among other things. What most investors in investment-grade debt focus on more than default risk is spread risk—that is, the effect on prices and returns from changes in spreads.

The price impact from spread changes is driven by two main factors: (1) the modified duration (price sensitivity with respect to changes in interest rates) of the bond and (2) the magnitude of the spread change. The effect on return to the bondholder depends on the holding period used for calculating the return.

The simplest example is that of a small, instantaneous change in the yield spread. In this case, the price impact (i.e., the percentage change in price, including accrued interest) can be approximated by

$$\%\Delta PV^{Full} \approx -\text{AnnModDur} \times \Delta\text{Spread},$$

where AnnModDur is the annual modified duration. The negative sign in this equation reflects the fact that because bond prices and yields move in opposite directions, lower spreads have a positive impact on bond prices and thus returns, whereas higher spreads have a negative impact on bond returns. Note that if the spread change is expressed in basis points, then the price impact will also be in basis points, whereas if the spread change is expressed as a decimal, the price impact will also be expressed as a decimal.

For larger spread changes (and thus larger yield changes), the impact of convexity needs to be incorporated into the approximation:

$$\%\Delta PV^{Full} \approx -(\text{AnnModDur} \times \Delta\text{Spread}) + \tfrac{1}{2}\text{AnnConvexity} \times (\Delta\text{Spread})^2.$$

In this case, one must be careful to ensure that convexity (denoted by AnnConvexity) is appropriately scaled to be consistent with the way the spread change is expressed. In general, for bonds without embedded options, one can scale convexity so that it has the same order of magnitude as the duration squared and then express the spread change as a decimal. For example, for a bond with duration of 5.0 and reported convexity of 0.235, one would re-scale convexity to 23.5 before applying the formula. For a 1% (i.e., 100 bps) increase in spread, the result would be

$$\%\Delta PV^{Full} = (-5.0 \times 0.01) + ½ \times 23.5 \times (0.01)^2 = -0.048825 \text{ or } -4.8825\%.$$

The price impact of instantaneous spread changes is illustrated in Exhibit 16 using two bonds from British Telecom, the UK telecommunications company. The bonds, denominated in British pounds, are priced to provide a certain spread over British government bonds (gilts) of a similar maturity. From the starting spread, in increments of 25 bps and for both wider and narrower spreads, the new price and actual return for each spread change are calculated. In addition, the exhibit shows the approximate returns with and without the convexity term. As can be seen, the approximation using only duration is reasonably accurate for small spread changes; for larger changes, however, the convexity term generally provides a meaningful improvement.

Exhibit 16 Impact of Duration on Price for a Given Change in Spread

Issuer: British Telecom, 5.75%, 12/07/2028

Price: £122.978	Modified Duration: 7.838	Spread to Gilt Curve: 150.7 b.p.
Accrued interest: 0.958	Convexity: 77.2	YTM (conv): 3.16

	Scenarios								
Spread Δ (b.p.)	−100	−75	−50	−25	0	25	50	75	100
Spread (b.p.)	50.7	75.7	100.7	125.7	150.7	175.7	200.7	225.7	250.7
Price (£)	131.62	129.12	126.68	124.29	122.98	119.69	117.47	115.30	113.18
Price + Accrued (£)	132.58	130.08	127.64	125.25	123.94	120.65	118.43	116.26	114.14
Price Δ (£)	8.64	6.14	3.70	1.31	0.00	−3.29	−5.51	−7.68	−9.80
Return (%)									
Actual	6.97%	4.96%	2.99%	1.06%	0.00%	−2.65%	−4.44%	−6.20%	−7.91%
Approx: Dur only	7.84%	5.88%	3.92%	1.96%	0.00%	−1.96%	−3.92%	−5.88%	−7.84%
Approx: Dur & Cvx	8.22%	6.10%	4.02%	1.98%	0.00%	−1.94%	−3.82%	−5.66%	−7.45%

Issuer: British Telecom, 3.625%, 21/11/2047

Price: £94.244	Modified Duration: 17.144	Spread to Gilt Curve: 210.8 b.p.
Accrued interest: 2.185	Convexity: 408.4	YTM (conv): 4.11

	Scenarios								
Spread Δ (b.p.)	−100	−75	−50	−25	0	25	50	75	100
Spread (b.p.)	110.8	135.8	160.8	185.8	210.8	235.8	260.8	285.8	310.8
Price (£)	111.28	106.38	101.77	97.41	93.24	89.41	85.75	82.30	79.04
Price + Accrued (£)	113.47	108.57	103.96	99.60	95.43	91.60	87.94	84.48	81.22
Price Δ (£)	18.04	13.14	8.53	4.17	0.00	−3.83	−7.49	−10.95	−14.21
Return (%)									
Actual	18.90%	13.77%	8.93%	4.37%	0.00%	−4.02%	−7.85%	−11.47%	−14.89%

Exhibit 16 (Continued)

	Scenarios								
Approx: Dur only	17.14%	12.86%	8.57%	4.29%	0.00%	−4.29%	−8.57%	−12.86%	−17.14%
Approx: Dur & Cvx	19.19%	14.01%	9.08%	4.41%	0.00%	−4.16%	−8.06%	−11.71%	−15.10%

Note: Settle date is 13 December 2018.
Source: Bloomberg Finance, L.P.

Note that the price change for a given spread change is higher for the longer-duration bond—in this case, the 2047 maturity British Telecom bond—than for the shorter-duration, 2028 maturity British Telecom bond. Longer-duration corporate bonds are referred to as having "higher spread sensitivity"; that is, their prices, and thus returns, are more volatile with respect to changes in spread. It is essentially the same concept as duration for any bond: The longer the duration of a bond, the greater the price volatility for a given change in interest rates/yields.

In addition, investors want to be compensated for the fact that the further time one is from a bond's maturity (i.e., the longer the bond), the greater the uncertainty about an issuer's future creditworthiness. Based on credit analysis, an investor might be confident that an issuer's risk of default is relatively low in the near term; however, looking many years into the future, the investor's uncertainty grows because of factors that are increasingly difficult, if not impossible, to forecast (e.g., poor management strategy or execution, technological obsolescence, natural or man-made disasters, corporate leveraging events). This increase in credit risk over time can be seen in Exhibit 17. Note that in this Standard & Poor's study, one-year default rates for the 2017 issuance pool are 0% for all rating categories of BB or higher. The three-year default rates for bonds issued in 2015 are materially higher, and the observed defaults include bonds originally rated up to BBB (i.e., low investment grade). The 5-year default rates for bonds issued in 2013 are higher than the 3-year default rates, and the defaults also include bonds initially rated as high as BBB. In addition to the risk of default rising over time, the data also show quite conclusively that the lower the credit rating, the higher the risk of default. Finally, note the very high risk of default for bonds rated CCC or lower over all time horizons. This is consistent with Exhibit 7 presented earlier, which showed significant three-year ratings variability ("migration"), with much of the migration to lower credit ratings (i.e., higher risk of default).

Exhibit 17 Default Rate by Rating Category (%) (Non-financials)

Credit Rating	1 Year (2017 pool)	3 Year (2015 pool)	10 Year (2013 pool)
AAA	0.00	0.00	0.00
AA	0.00	0.00	0.00
A	0.00	0.00	0.00
BBB	0.00	0.08	0.27
BB	0.10	2.46	3.33

(continued)

Exhibit 17 (Continued)

Credit Rating	1 Year (2017 pool)	3 Year (2015 pool)	10 Year (2013 pool)
B	0.95	10.11	12.90
CCC/C	27.15	41.43	44.70

Source: Based on data from S&P Global Ratings, "2017 Annual Global Corporate Default Study and Rating Transitions" (5 April 2018).

EXAMPLE 9

Price Impact

Calculate the price impact on a 10-year corporate bond with a 4.75% coupon priced at 100, with an instantaneous 50 bps widening in spread due to the issuer's announcement that it was adding substantial debt to finance an acquisition (which resulted in a two-notch downgrade by the rating agencies). The bond has a modified duration of 7.9, and its convexity is 74.9.

Solution:

The impact from the 50 bps spread widening is:

$$\text{Price impact} \approx -(\text{AnnModDur} \times \Delta\text{Spread}) + ½\ \text{AnnConvexity} \times (\Delta\text{Spread})^2$$
$$= -(0.0050 \times 7.9) + (0.5 \times 74.9) \times (0.0050)^2$$
$$= -0.0386, \text{ or } -3.86\%.$$

Because yields and bond prices move in opposite directions, the wider spread caused the bond price to fall. Using a bond-pricing calculator, the exact return is −3.85%, so this approximation was very accurate.

In summary, spread changes can have a significant impact on the price and performance of credit-risky bonds over a given holding period, and the higher the modified duration of the bond(s), the greater the price impact from changes in spread. Wider spreads hurt bond performance, whereas narrower spreads help bond performance. For bond investors who actively manage their portfolios (i.e., don't just buy bonds and hold them to maturity), forecasting spread changes and expected credit losses on both individual bonds and their broader portfolios is an important strategy for enhancing investment performance.

7 HIGH-YIELD, SOVEREIGN, AND NON-SOVEREIGN CREDIT ANALYSIS

j explain special considerations when evaluating the credit of high-yield, sovereign, and non-sovereign government debt issuers and issues

Thus far, we have focused primarily on basic principles of credit analysis and investing with emphasis on higher-quality, investment-grade corporate bonds. Although many of these principles are applicable to other credit-risky segments of the bond market, some differences in credit analysis need to be considered. This section focuses on special considerations in evaluating the credit of debt issuers from the following three market segments: high-yield corporate bonds, sovereign bonds, and non-sovereign government bonds.

7.1 High Yield

Recall that high-yield, or non-investment-grade, corporate bonds are those rated below Baa3/BBB– by the major rating agencies. These bonds are sometimes referred to as "junk bonds" because of the higher risk inherent in their weak balance sheets and/or poor or less-proven business prospects.

Companies are rated below investment grade for many reasons, including the following:

- Highly leveraged capital structure
- Weak or limited operating history
- Limited or negative free cash flow
- Highly cyclical business
- Poor management
- Risky financial policies
- Lack of scale and/or competitive advantages
- Large off-balance-sheet liabilities
- Declining industry (e.g., newspaper publishing)

Companies with weak balance sheets and/or business profiles have lower margin for error and greater risk of default relative to higher-quality investment-grade names. And the higher risk of default means more attention must be paid to recovery analysis (or loss severity, in the event of default). Consequently, high-yield analysis typically is more in-depth than investment-grade analysis and thus has special considerations. This includes the following:

- Greater focus on issuer liquidity and cash flow
- Detailed financial projections
- Detailed understanding and analysis of the debt structure
- Understanding of an issuer's corporate structure
- Covenants
- Equity-like approach to high-yield analysis

Liquidity

Liquidity—that is, having cash and/or the ability to generate or raise cash—is important to all issuers. It is absolutely critical for high-yield companies. Investment-grade companies typically have substantial cash on their balance sheets, generate a lot of cash from operations relative to their debt (or else they wouldn't be investment grade!), and/or are presumed to have alternate sources of liquidity, such as bank lines and commercial paper. For these reasons, investment-grade companies can more easily roll over (refinance) maturing debt. On the other hand, high-yield companies may not have those options available. For example, there is no high-yield commercial paper market, and bank credit facilities often carry tighter restrictions for high-yield companies. Both bad company-specific news and difficult financial market conditions can lead to high-yield companies being unable to access the debt markets. And although the vast majority of investment-grade corporate debt issuers have publicly traded equity and can thus use that equity as a financing option, many high-yield companies are privately held and thus don't have access to public equity markets.

Thus, issuer liquidity is a key focus in high-yield analysis. Sources of liquidity, from strongest to weakest, are the following:

1 Cash on the balance sheet
2 Working capital
3 Operating cash flow
4 Bank credit facilities
5 Equity issuance
6 Asset sales

Cash on the balance sheet is easy to see and self-evident as a source for repaying debt, although a portion of it may be 'trapped' overseas for certain tax, business, accounting, or regulatory reasons and thus not easily accessible. As mentioned earlier, working capital can be a large source or use of liquidity, depending on its amount, its use in a company's cash-conversion cycle, and its role in a company's operations. Operating cash flow is a ready source of liquidity as sales turn to receivables, which turn to cash over a fairly short time period. Bank lines, or credit facilities, can be an important source of liquidity, though there may be some covenants relating to the use of the bank lines that are crucial to know and will be covered a little later. Equity issuance may not be a reliable source of liquidity because an issuer is private or because of poor market conditions if a company does have publicly traded equity. Asset sales are the least reliable source of liquidity because both the potential value and the actual time of closing can be highly uncertain.

The amount of these liquidity sources should be compared with the amount and timing of upcoming debt maturities. A large amount of debt coming due in the next 6–12 months alongside low sources of liquidity will be a warning flag for bond investors and could push an issuer into default because investors may choose not to buy new bonds intended to pay off the existing debt. Insufficient liquidity—that is, running out of cash or no longer having access to external financing to refinance or pay off existing debt—is the principal reason issuers default. Although liquidity is important for industrial companies, it is an absolute necessity for financial firms, as seen in the case of Lehman Brothers and other troubled firms during the global financial crisis of 2008. Financial institutions are highly levered and often highly dependent on funding longer-term assets with short-term liabilities.

Financial Projections

Because high-yield companies have less room for error, it's important to forecast, or project, future earnings and cash flow out several years, perhaps including several scenarios, to assess whether the issuer's credit profile is stable, improving, or declining and thus whether it needs other sources of liquidity or is at risk of default. Ongoing capital expenditures and working capital changes should be incorporated as well. Special emphasis should be given to realistic "stress" scenarios that could expose a borrower's vulnerabilities.

Debt Structure

High-yield companies tend to have many layers of debt in their capital structures, with varying levels of seniority and, therefore, different potential recovery rates in the event of default. (Recall the historical table of default recovery rates based on seniority in Exhibit 2.) A high-yield issuer will often have at least some of the following types of obligations in its debt structure:

- (Secured) Bank debt
- Second lien debt
- Senior unsecured debt
- Subordinated debt, which may include convertible bonds
- Preferred stock

The lower the ranking in the debt structure, the lower the credit rating and the lower the expected recovery in the event of default. In exchange for these associated higher risks, investors will normally demand higher yields.

As discussed earlier, a standard leverage calculation used by credit analysts is debt/EBITDA and is quoted as a multiple (e.g., "5.2x levered"). For an issuer with several layers of debt with different expected recovery rates, high-yield analysts should calculate leverage at each level of the debt structure. Example 10 shows calculations of gross leverage, as measured by debt/EBITDA, at each level of the debt structure and net leverage for the entire debt structure. Gross leverage calculations do not adjust debt for cash on hand. Net leverage adjusts debt by subtracting cash from total debt.

EXAMPLE 10

Debt Structure and Leverage

Hexion Inc. is a specialty chemical company. It has a complicated, high-yield debt structure, consisting of first lien debt (loans and bonds), secured bonds, second lien bonds, and senior unsecured debt, due to a series of mergers as well as a leveraged buyout in 2005. Exhibit 18 is a simplified depiction of the company's debt structure, as well as some key credit-related statistics.

Exhibit 18 Hexion Inc. Debt and Leverage Structure as of Year-End 2017

Financial Information ($ millions)	
Cash	$115
Total debt	$3,668
Net debt	$3,553
Interest expense	$329
EBITDA	$365

Debt Structure ($ millions)	
First lien debt (loans and bonds)	$2,607
Secured bonds	$225
Second lien bonds	$574

(continued)

Exhibit 18 (Continued)

Debt Structure ($ millions)	
Senior unsecured bonds	$263
TOTAL DEBT	$3,669

Source: Company filings, Loomis, Sayles & Company.

Using the information provided, address the following:

1 Calculate gross leverage, as measured by debt/EBITDA, through each level of debt, including total debt.

2 Calculate the net leverage, as measured by (Debt – Cash)/EBITDA, for the total debt structure.

3 Why might Hexion have so much secured debt relative to unsecured debt (both senior and subordinated)? (*Note*: This question draws on concepts from earlier sections.)

Solutions to 1 and 2:

	Gross Leverage (Debt/EBITDA)	Net Leverage (Debt – Cash)/ EBITDA
Secured debt leverage		
(First lien + Secured debt)/EBITDA		
(2,607 + 225)/365	7.8x	
Second lien leverage		
(First lien + Secured debt + Second lien debt)/EBITDA		
(2,607 + 225 + 574)/365	9.3x	
Total leverage (includes unsecured)		
(Total debt/EBITDA)		
3,669/365	10.1x	
Net leverage (leverage net of cash through entire debt structure)		
(Total debt – Cash)/EBITDA		9.7x

Solution to 3:

Hexion might have that much secured debt because (1) it was less expensive than issuing additional unsecured debt on which investors would have demanded a higher yield and/or (2) given the riskiness of the business (chemicals are a cyclical business), the high leverage of the business model, and the riskiness of the balance sheet (lots of debt from a leveraged buyout), investors would only be willing to lend the company money on a secured basis.

High-yield companies that have a lot of secured debt (typically bank debt) relative to unsecured debt are said to have a "top-heavy" capital structure. With this structure, there is less capacity to take on more bank debt in the event of financial stress. Along with the often more stringent covenants associated with bank debt and its generally shorter maturity compared with other types of debt, this means that these issuers are more susceptible to default, as well as to lower recovery for the various less secured creditors.

Corporate Structure

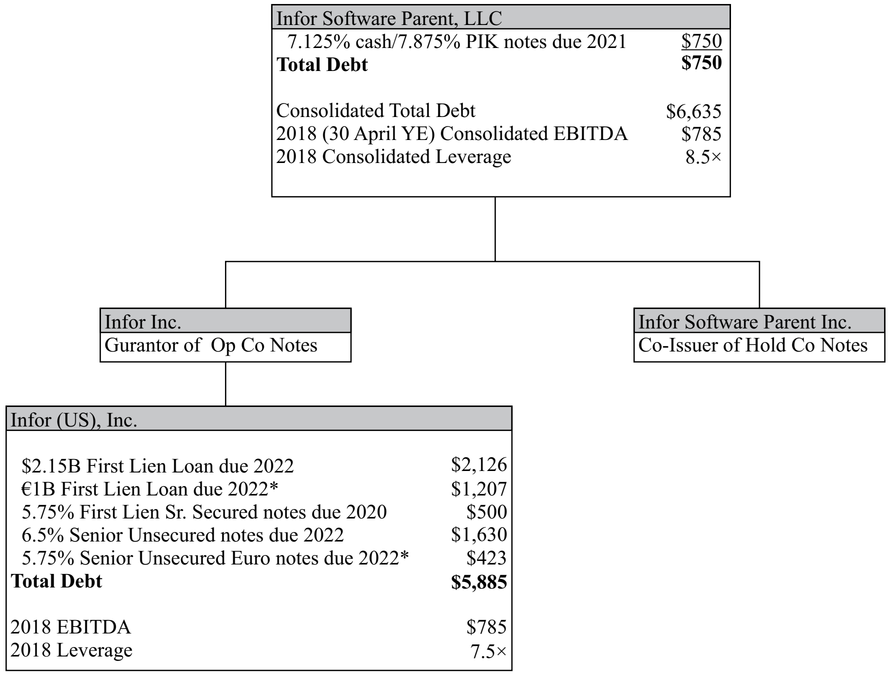
Many debt-issuing corporations, including high-yield companies, utilize a holding company structure with a parent and several operating subsidiaries. Knowing where an issuer's debt resides (parent versus subsidiaries) and how cash can move from subsidiary to parent ("upstream") and vice versa ("downstream") are critical to the analysis of high-yield issuers.

In a holding company structure, the parent owns stock in its subsidiaries. Typically, the parent doesn't generate much of its own earnings or cash flow but instead receives dividends from its subsidiaries. The subsidiaries' dividends are generally paid out of earnings after they satisfy all their other obligations, such as debt payments. To the extent that their earnings and cash flow are weak, subsidiaries may be limited in their ability to pay dividends to the parent. Moreover, subsidiaries that carry a lot of their own debt may have restrictions or limitations on how much cash they can provide to the parent via dividends or in another way, such as through an intercompany loan. These restrictions and limitations on cash moving between parent and subsidiaries can have a major impact on their respective abilities to meet their debt obligations. The parent's reliance on cash flow from its subsidiaries means the parent's debt is structurally subordinated to the subsidiaries' debt and thus will usually have a lower recovery rating in default.

For companies with very complex holding companies, there may also be one or more intermediate holding companies, each carrying their own debt, and in some cases, they may not own 100% of the subsidiaries' stock. Often times a default in one subsidiary may not trigger a cross-default. This structure is sometimes seen in high-yield companies that have been put together through many mergers and acquisitions or that were part of a leveraged buyout.

Exhibit 19 shows the capital structure of Infor, Inc. (Infor), a high-yield software and services company highlighted earlier as an example of the credit rating agency notching process. Infor's capital structure consists of a parent company that has debt—in this case, convertible senior notes—as well as a subsidiary with multiple layers of outstanding debt by seniority.

Exhibit 19 Infor's Capital Structure



*Conversion from euro to US dollar.
Source: Company filing, Loomis, Sayles & Company.

Thus, high-yield investors should analyze and understand an issuer's corporate structure, including the distribution of debt between the parent and its subsidiaries. Leverage ratios should be calculated at each of the debt-issuing entities, as well as on a consolidated basis.

Also important is that although the debt of an operating subsidiary may be "closer to" and better secured by particular assets of the subsidiary, the credit quality of a parent company might still be higher. The parent company could, while being less directly secured by any particular assets, still benefit from the diversity and availability of all the cash flows in the consolidated system. In short, credit quality is not simply an automatic analysis of debt provisions and liens.

Covenant Analysis

As discussed earlier, analysis of covenants is very important for all bonds. It is especially important for high-yield credits because of their reduced margin of safety. Key covenants for high-yield issuers may include the following:

- Change of control put
- Restricted payments
- Limitations on liens and additional indebtedness
- Restricted versus unrestricted subsidiaries

Under the **change of control put**, in the event of an acquisition (a "change of control"), bondholders have the right to require the issuer to buy back their debt (a "put option"), often at par or at some small premium to par value. This covenant is intended to protect creditors from being exposed to a weaker, more indebted borrower as a result of acquisition. For investment-grade issuers, this covenant typically has a two-pronged test: acquisition of the borrower and a consequent downgrade to a high-yield rating.

The **restricted payments** covenant is meant to protect creditors by limiting how much cash can be paid out to shareholders over time. The restricted payments "basket" is typically sized relative to an issuer's cash flow and debt outstanding—or is being raised—and is an amount that can grow with retained earnings or cash flow, giving management more flexibility to make payouts.

The **limitations on liens** covenant is meant to put limits on how much secured debt an issuer can have. This covenant is important to unsecured creditors who are structurally subordinated to secured creditors; the higher the amount of debt that is layered ahead of them, the less they stand to recover in the event of default.

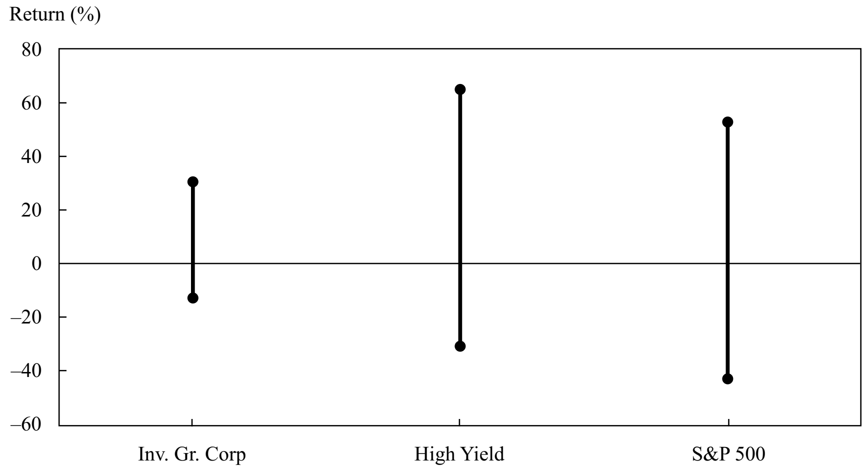
With regard to restricted versus unrestricted subsidiaries, issuers may classify certain of their subsidiaries as restricted and others as unrestricted as it pertains to offering guarantees for their holding company debt. These subsidiary guarantees can be very useful to holding company creditors because they put their debt on equal standing (pari passu) with debt at the subsidiaries instead of with structurally subordinated debt. Restricted subsidiaries should be thought of as those that are designated to help service parent-level debt, typically through guarantees. They tend to be an issuer's larger subsidiaries and have significant assets, such as plants and other facilities, and/or cash flow. There may be tax or legal (e.g., country of domicile) reasons why certain subsidiaries are restricted while others are not. Analysts should carefully read the definitions of restricted versus unrestricted subsidiaries in the indenture because sometimes the language is so loosely written that the company can reclassify subsidiaries from one type to another with a simple vote by a board of directors or trustees.

For high-yield investors, it is also important to know what covenants are in an issuer's bank credit agreements. These agreements are typically filed with the securities commission in the country where the loan document was drafted. Bank covenants can be more restrictive than bond covenants and may include so-called **maintenance covenants**, such as leverage tests, whereby the ratio of, say, debt/EBITDA may not exceed "x" times. In the event a covenant is breached, the bank is likely to block further loans under the agreement until the covenant is cured. If not cured, the bank may accelerate full payment of the facility, triggering a default.

Equity-like approach to high-yield analysis

High-yield bonds are sometimes thought of as a "hybrid" between higher-quality bonds, such as investment-grade corporate debt, and equity securities. Their more volatile price and spread movements are less influenced by interest rate changes than are higher-quality bonds, and they show greater correlation with movements in equity markets. Indeed, as shown in Exhibit 20, historical returns on high-yield bonds and the standard deviation of those returns fall somewhere between investment-grade bonds and equities.

Exhibit 20 US Trailing 12-Month Returns by Asset Class, 31 December 1988–30 September 2018



	Inv. Gr. Corp	High Yield	S&P 500
Avg. Return:	7.00%	8.86%	11.39%
Max. Return:	31.07%	64.95%	53.62%
Min. Return:	–13.82%	–31.23%	–43.32%
Avg. Volatility:	4.71%	6.79%	12.91%

Sources: Bloomberg Barclays Indices; Haver Analytics; and Loomis, Sayles & Company.

Consequently, an equity market-like approach to analyzing a high-yield issuer can be useful. One approach is to calculate an issuer's enterprise value. Enterprise value (EV) is usually calculated by adding equity market capitalization and total debt and then subtracting excess cash. Enterprise value is a measure of what a business is worth (before any takeover premium) because an acquirer of the company would have to either pay off or assume the debt and it would receive the acquired company's cash.

Bond investors like using EV because it shows the amount of equity "cushion" beneath the debt. It can also give a sense of (1) how much more leverage management might attempt to put on a company in an effort to increase equity returns or (2) how likely—and how expensive—a credit-damaging leveraged buyout might be. Similar to how stock investors look at equity multiples, bond investors may calculate and compare EV/EBITDA and debt/EBITDA across several issuers as part of their analysis. Narrow differences between EV/EBITDA and debt/EBITDA for a given issuer indicate a small equity cushion and, therefore, potentially higher risk for bond investors.

7.2 Sovereign Debt

Governments around the world issue debt to help finance their general operations, including such current expenses as wages for government employees, and investments in such long-term assets as infrastructure and education. Government bonds in developed countries have traditionally been viewed as the default risk-free rate off of which all other credits are priced. Fiscal challenges in developed countries exacerbated by the 2008 global financial crisis and the 2011–2012 eurozone crisis, however, have called into question the notion of a "risk-free rate," even for some of the highest-quality government borrowers. As their capital markets have developed, an increasing number of sovereign governments have been able to issue debt in foreign markets (generally denominated in a currency other than that of the sovereign government, often the US dollar or euro) as well as debt in the domestic market (issued in the sovereign government's own currency). Generally, sovereign governments with weak currencies can access foreign debt markets only by issuing bonds in foreign currencies that are viewed to be safer stores of value. Debt issued in the domestic market is somewhat easier to service because the debt is typically denominated in the country's own currency, subject to its own laws, and money can be printed to service the sovereign government's domestic debt. Twenty years ago, many emerging market countries could only issue debt in foreign markets because a domestic market did not exist. Today, many are able to issue debt domestically and have successfully built yield curves of domestic bonds across the maturity spectrum. All sovereign governments are best able to service foreign and domestic debt if they run "twin surpluses"—that is, a government budget surplus as well as a current account surplus.

Despite ongoing financial globalization and the development of domestic bond markets, sovereign government defaults occur. Defaults are often precipitated by such events as war, political upheaval, major currency devaluation, a sharp deterioration in trade, or dramatic price declines in a country's key commodity exports. Default risks for some developed countries escalated after 2009 as government revenues dropped precipitously following the global financial crisis of 2008, expenditures surged, and financial markets focused on the long-term sustainability of public finances, given aging populations and rising social security needs. Some of the weaker and more highly indebted members of the eurozone became unable to access the debt markets at economical rates and had to seek loans from the International Monetary Fund (IMF) and the European Union. These weaker governments had previously been able to borrow at much lower rates because of their membership in the European Union and adoption of the euro. Intra-eurozone yield spreads widened and countries were shut out of markets, however, as the global financial crisis exacted a high toll on their public finances and, in some cases, their banking systems, which became contingent liabilities for the sovereigns. In Ireland, the government guaranteed most bank liabilities, undermining the country's own fiscal metrics.

Like corporate analysis, sovereign credit analysis is based on a combination of qualitative and quantitative factors. Ultimately, the two key issues for sovereign analysis are: (1) a government's ability to pay and (2) its willingness to pay. Willingness to pay is important because of the principle of sovereign immunity, where investors are generally unable to force a sovereign to pay its debts. Sovereign immunity prevents governments from being sued. To date, global initiatives aimed at creating a mechanism for orderly sovereign restructurings and defaults have not found traction.

To illustrate the most important considerations in sovereign credit analysis, we present a basic framework for evaluating sovereign credit and assigning sovereign debt ratings (this outline was developed from the detailed exposition of Standard & Poor's methodology given in S&P Global Ratings, "Sovereign Rating Methodology" [18 December 2017]). The framework highlights the specific characteristics analysts should expect in a high-quality sovereign credit. Some of these are self-explanatory (e.g., absence of corruption). For others, a brief rationale and/or range of values is included to clarify interpretation. Most, but not all, of these items are included in rating agency Standard & Poor's methodology.

Institutional and economic profile

- *Institutional assessment*
  - Institutions' ability to deliver sound public finances and balanced economic growth.
  - Effectiveness and predictability of policymaking institutions.
  - Track record of managing previous political, economic, and/or financial crises.
  - Ability and willingness to implement reforms to address fiscal challenges.

- Transparent and accountable institutions with low perceived level of corruption.
- Independence of statistical offices and media.
- Checks and balances between institutions.
- Unbiased enforcement of contracts and respect for rule of law and property rights.
- Debt repayment culture.
- Potential external and domestic security risks.

■ *Economic assessment*

- Income per capita: More prosperous countries generally have a broader and deeper tax base with which to support debt.
- Trend growth prospects: Creditworthiness is supported by sustainable and durable trend growth across business cycles.
- Diversity and stability of growth: Sovereigns exposed to economic concentration are more vulnerable. A narrow economy tends to show higher volatility in growth and can impair a government's balance sheet.

Flexibility and performance profile

■ *External assessment*

- Status of currency: Sovereigns that control a reserve currency or a very actively traded currency are able to use their own currency in many international transactions and are less vulnerable to adverse shifts in global investor portfolios.
- External liquidity: Countries with a substantial supply of foreign currency (foreign exchange reserves plus current account receipts) relative to projected funding needs in foreign currency (current account payments plus debt maturities) are less vulnerable to interruption of external liquidity.
- External debt: Countries with low foreign debt relative to current account receipts are better able to service their foreign debt. This is similar to a coverage ratio for a corporation.

■ *Fiscal assessment*

- Fiscal performance and flexibility: Trend change in net general government debt as a percentage of GDP. Stable or declining debt as a percentage of GDP indicates a strong credit; a rising ratio can prove unsustainable and is, therefore, a sign of diminishing creditworthiness.
- Long-term fiscal trends: Perceived willingness and ability to increase revenue or cut expenditure to ensure debt service.
- Debt burden and structure: Net general government debt of less than 30% is good; more than 100% is poor. General government interest expense as a percentage of revenue: Less than 5% is good; greater than 15%is poor.
- Ability to access funding, manage both the amortization profile and contingent liabilities arising from the financial sector, public enterprises, and guarantees.

■ *Monetary assessment*

- Ability to use monetary policy tailored to domestic economic objectives (e.g., growth) to address imbalances or shocks.

- Exchange rate regime: Sovereigns with a reserve currency have the most flexibility. A freely floating currency allows maximum effectiveness for monetary policy. A fixed-rate regime limits effectiveness and flexibility. A hard peg, such as a currency board or monetary union, affords no independent monetary policy.
- Credibility of monetary policy: Measured by track record of low and stable inflation. Credible monetary policy is supported by an operationally and legally independent central bank with a clear mandate. The central bank's ability to be a lender of last resort to the financial system also enhances stability.
- Confidence in the central bank provides a foundation for confidence in the currency as a store of value and for the central bank to effectively manage policy through a crisis.
- Most effective policy transmission occurs in systems with sound banking systems and well-developed domestic capital markets, including active money market and corporate bond markets, such that policymakers credibly enact policy relying on market-based policy tools (e.g., open market operations) over administrative policy tools (e.g., reserve requirements).

In light of a sovereign government's various powers—taxation, regulation, monetary policy, and ultimately, the sovereign's ability to "print money" to repay debt—within its own economy, it is virtually always at least as good a credit in its domestic currency as it is in foreign currency. Thus, credit rating agencies often distinguish between domestic and foreign bonds, with domestic bond ratings sometimes one notch higher. Of course, if a sovereign government were to rely heavily on printing money to repay debt, it would fuel high inflation or hyperinflation and increase default risk on domestic debt as well.

EXAMPLE 11

Sovereign Debt

Exhibit 21 shows several key sovereign statistics for Portugal.

Exhibit 21 Key Sovereign Statistics for Portugal

€ (billions), except where noted	2006	2008	2010	2012	2014	2015	2016	2017
Nominal GDP	160.3	171.2	172.6	168.4	173.1	179.8	186.5	194.6
Population (millions)	10.6	10.6	10.6	10.5	10.4	10.3	10.3	10.3
Unemployment (%)	8.6	8.5	12	15.6	13.9	12.4	11.1	8.9
Exports as share GDP (%)	22.2	22.6	21.3	26.8	27.8	27.6	26.8	28.3
Current account as share GDP (%)	−10.7	−12.6	−10	−2.1	0.2	0.3	0.7	0.6
Government revenues	64.8	70.7	71.8	72.2	77.2	78.8	79.9	83.1
Government expenditures	71.4	77.1	88.7	81.7	89.6	86.7	83.5	88.9
Budget balance (surplus/deficit)	−6.5	−6.4	−16.9	−9.5	−12.4	−7.9	−3.6	−5.8
Government interest payments	4.2	5.3	5.2	8.2	8.5	8.2	7.8	7.4
Primary balance (surplus/deficit)	−2.2	−1.1	−11.7	−1.3	−3.9	0.3	4.2	1.6

(continued)

Exhibit 21 (Continued)

€ (billions), except where noted	2006	2008	2010	2012	2014	2015	2016	2017
Government debt	102.4	123.1	161.3	212.6	226	231.5	241	242.8
Interest rate on new debt (%)	3.9	4.5	5.4	3.4	3.8	2.4	3.2	3.1

Sources: Haver Analytics, Eurostat, and Instituto Nacional de Estatística (Portugal).

1 Calculate the government debt/GDP for Portugal for the years 2014–2017 as well as for the years 2006, 2008, 2010, and 2012.

2 Calculate GDP/capita for the same periods.

3 Based on those calculations, as well as other data from Exhibit 21, what can you say about Portugal's credit trend?

Solutions to 1 and 2:

	2006	2008	2010	2012	2014	2015	2016	2017
Gross government debt/GDP	64%	72%	93%	126%	131%	129%	129%	125%
GDP/capita	15,123	16,151	16,283	16,038	16,644	17,456	18,107	18,893

Solution to 3:

The credit trend is stabilizing. Government debt/GDP is declining ever so slightly after peaking in 2014. The government is running a modest budget deficit with a primary balance that is in surplus for the past three years. Portugal is running a very small current account surplus, reducing its reliance on external funding, and has increased its exports as a share of GDP. Unemployment, while still fairly high, has fallen in the past several years. Interest payments on government debt have started to decline, both as a percentage of GDP and in absolute terms. The interest rate on new government debt has stabilized, perhaps benefitting from the European Central Bank's quantitative easing policies. Taken together, there are strong indications that the Portuguese government's credit situation has stabilized and may be expected to improve further if current trends are sustained.

7.3 Non-Sovereign Government Debt

Sovereigns are the largest issuers of government debt, but non-sovereign—sometimes called sub-sovereign or local—governments and the quasi-government entities that are created by governments issue bonds as well. The non-sovereign or local governments include governments of states, provinces, regions, and cities. For example, the City of Tokyo (Tokyo Metropolitan Government) has debt outstanding, as does the Lombardy region in Italy, the City of Buenos Aires in Argentina, and the State of California in the United States. Local government bonds may be referred to as municipal bonds.

However, when people talk about municipal bonds, they are usually referring to US municipal bonds, which represent one of the largest bond markets. As of year-end 2017, the US municipal bond market was approximately $3.9 trillion in size, roughly 9% of the total US bond market (Securities Industry and Financial Markets Association [SIFMA], "Outstanding U.S. Bond Market Data," as of 2Q 2018). The US municipal bond market is composed of both tax-exempt and, to a lesser extent, taxable bonds issued by state and city governments and their agencies. Municipal borrowers may also issue bonds on behalf of private entities, such as non-profit colleges or hospitals.

Historically, for any given rating category, these bonds have much lower default rates than corporate bonds with the same ratings. For example, according to Moody's Investors Service ("US Municipal Bond Defaults and Recoveries, 1970–2017"), the 10-year average cumulative default rate from 1970 through 2017 was 0.17% for municipal bonds, compared with a 10.24% 10-year average cumulative default rate for all corporate debt.

The majority of local government bonds, including municipal bonds, are either general obligation bonds or revenue bonds. General obligation (GO) bonds are unsecured bonds issued with the full faith and credit of the issuing non-sovereign government. These bonds are supported by the taxing authority of the issuer. Revenue bonds are issued for specific project financing (e.g., financing for a new sewer system, a toll road, bridge, hospital, or sports arena).

The credit analysis of GO bonds has some similarities to sovereign debt analysis (e.g., the ability to levy and collect taxes and fees to help service debt) but also some differences. For example, almost without exception, US municipalities must balance their operating budgets (i.e., exclusive of long-term capital projects) annually. Non-sovereign governments are unable to use monetary policy the way many sovereigns can.

The economic analysis of non-sovereign government GO bonds, including US municipal bonds, focuses on employment, per capita income (and changes in it over time), per capita debt (and changes in it over time), the tax base (depth, breadth, diversification, stability, etc.), demographics, and net population growth, as well as an analysis of whether the area represented by the non-sovereign government has the infrastructure and location to attract and support new jobs. Analysis should look at the volatility and variability of revenues during times of both economic strength and weakness. An overreliance on one or two types of tax revenue—particularly a volatile one, such as capital gains taxes or sales taxes—can signal increased credit risk. Pensions and other post-retirement obligations may not show up directly on the non-sovereign government's balance sheet, and many of these entities have underfunded pensions that need to be addressed. Adding the unfunded pension and post-retirement obligations to the debt reveals a more realistic picture of the issuer's debt and longer-term obligations. The relative ease or difficulty in managing the annual budgeting process and the government's ability to operate consistently within its budget are also important credit analysis considerations.

Disclosure by non-sovereign governments varies widely, with some of the smaller issuers providing limited financial information. Reporting requirements are inconsistent, so the financial reports may not be available for six months or more after the closing of a reporting period.

Exhibit 22 compares several key debt statistics from two of the larger states in the United States: Illinois and Texas. Illinois has the lowest credit ratings of any of the states, whereas Texas has one of the highest. Note the higher debt burden (and lower ranking) across several measures: Total debt, Debt/Capita, Debt/Personal income, and debt as a percentage of state GDP. When including net pension liabilities of government employees and retirees, the debt burdens are even greater, especially in the case of Illinois. What is not shown here is that Illinois also has a higher tax burden and greater difficulty balancing its budget on an annual basis than Texas does.

Exhibit 22 Municipal Debt Comparison: Illinois vs. Texas

	Illinois	Texas
Ratings:		
Moody's	Baa3	Aaa
S&P	BBB–	AAA

(continued)

Exhibit 22 (Continued)

	Illinois	Texas
Fitch	BBB	AAA
Unemployment rate (%)*	4.20	3.70
Median household income ($)**	$61,229	$57,051
Debt burden, net ($/rank)***		
Total (millions)	37,374 (5)	11,603 (13)
Per capita	2,919 (6)	410 (42)
As a percentage of 2016 personal income	5.60 (5)	0.90 (42)
As a percentage of 2016 GDP	4.70 (6)	0.73 (42)
ANPL, net ($/rank)****		
Total (millions)	250,136 (1)	140,253 (3)
Per capita	19,539 (1)	4,955 (19)
As a percentage of 2017 personal income	37.00 (1)	10.60 (19)
As a percentage of 2017 GDP	30.50 (1)	8.30 (20)

* *Source:* Bureau of Labor Statistics, data as of October 2018.
** *Source:* US Census Bureau, data as of 2017.
*** *Source:* Moody's Investors Service, Inc., debt data as of 2017.
**** *Source:* Moody's Investors Service, Inc., adjusted net pension liability data as of 2017.

Revenue bonds, which are issued to finance a specific project, have a higher degree of risk than GO bonds because they are dependent on a single source of revenue. The analysis of these bonds is a combination of an analysis of the project and the finances around the particular project. The project analysis focuses on the need and projected utilization of the project, as well as on the economic base supporting the project. The financial analysis has some similarities to the analysis of a corporate bond in that it is focused on operating results, cash flow, liquidity, capital structure, and the ability to service and repay the debt. A key credit measure for revenue-backed non-sovereign government bonds is the debt-service-coverage (DSC) ratio, which measures how much revenue is available to cover debt payments (principal and interest) after operating expenses. Many revenue bonds have a minimum DSC ratio covenant; the higher the DSC ratio, the stronger the creditworthiness.

SUMMARY

We introduced basic principles of credit analysis. We described the importance of the credit markets and credit and credit-related risks. We discussed the role and importance of credit ratings and the methodology associated with assigning ratings, as well as the risks of relying on credit ratings. We covered the key components of credit analysis and the financial measure used to help assess creditworthiness.

We also discussed risk versus return when investing in credit and how spread changes affect holding period returns. In addition, we addressed the special considerations to take into account when doing credit analysis of high-yield companies, sovereign borrowers, and non-sovereign government bonds.

- Credit risk is the risk of loss resulting from the borrower failing to make full and timely payments of interest and/or principal.
- The key components of credit risk are risk of default and loss severity in the event of default. The product of the two is expected loss. Investors in higher-quality bonds tend not to focus on loss severity because default risk for those securities is low.
- Loss severity equals (1 – Recovery rate).
- Credit-related risks include downgrade risk (also called credit migration risk) and market liquidity risk. Either of these can cause yield spreads—yield premiums—to rise and bond prices to fall.
- Downgrade risk refers to a decline in an issuer's creditworthiness. Downgrades will cause its bonds to trade with wider yield spreads and thus lower prices.
- Market liquidity risk refers to a widening of the bid–ask spread on an issuer's bonds. Lower-quality bonds tend to have greater market liquidity risk than higher-quality bonds, and during times of market or financial stress, market liquidity risk rises.
- The composition of an issuer's debt and equity is referred to as its "capital structure." Debt ranks ahead of all types of equity with respect to priority of payment, and within the debt component of the capital structure, there can be varying levels of seniority.
- With respect to priority of claims, secured debt ranks ahead of unsecured debt, and within unsecured debt, senior debt ranks ahead of subordinated debt. In the typical case, all of an issuer's bonds have the same probability of default due to cross-default provisions in most indentures. Higher priority of claim implies higher recovery rate—lower loss severity—in the event of default.
- For issuers with more complex corporate structures—for example, a parent holding company that has operating subsidiaries—debt at the holding company is structurally subordinated to the subsidiary debt, although the possibility of more diverse assets and earnings streams from other sources could still result in the parent having higher effective credit quality than a particular subsidiary.
- Recovery rates can vary greatly by issuer and industry. They are influenced by the composition of an issuer's capital structure, where in the economic and credit cycle the default occurred, and what the market's view of the future prospects are for the issuer and its industry.
- The priority of claims in bankruptcy is not always absolute. It can be influenced by several factors, including some leeway accorded to bankruptcy judges, government involvement, or a desire on the part of the more senior creditors to settle with the more junior creditors and allow the issuer to emerge from bankruptcy as a going concern, rather than risking smaller and delayed recovery in the event of a liquidation of the borrower.
- Credit rating agencies, such as Moody's, Standard & Poor's, and Fitch, play a central role in the credit markets. Nearly every bond issued in the broad debt markets carries credit ratings, which are opinions about a bond issue's creditworthiness. Credit ratings enable investors to compare the credit risk of debt issues and issuers within a given industry, across industries, and across geographic markets.

- Bonds rated Aaa to Baa3 by Moody's and AAA to BBB− by Standard & Poor's (S&P) and/or Fitch (higher to lower) are referred to as "investment grade." Bonds rated lower than that—Ba1 or lower by Moody's and BB+ or lower by S&P and/or Fitch—are referred to as "below investment grade" or "speculative grade." Below-investment-grade bonds are also called "high-yield" or "junk" bonds.
- The rating agencies rate both issuers and issues. Issuer ratings are meant to address an issuer's overall creditworthiness—its risk of default. Ratings for issues incorporate such factors as their rankings in the capital structure.
- The rating agencies will notch issue ratings up or down to account for such factors as capital structure ranking for secured or subordinated bonds, reflecting different recovery rates in the event of default. Ratings may also be notched due to structural subordination.
- Rating agencies incorporate ESG factors into their ratings of firms. Some have launched a set of ratings that aim to measure a company's attitudes, practices, and advances related to ESG. They identify and track leaders and laggards in the space. Companies are evaluated according to their exposure to ESG risks and how well they manage those risks relative to peers.
- There are risks in relying too much on credit agency ratings. Creditworthiness may change over time, and initial/current ratings do not necessarily reflect the creditworthiness of an issuer or bond over an investor's holding period. Valuations often adjust before ratings change, and the notching process may not adequately reflect the price decline of a bond that is lower ranked in the capital structure. Because ratings primarily reflect the probability of default but not necessarily the severity of loss given default, bonds with the same rating may have significantly different expected losses (default probability times loss severity). And like analysts, credit rating agencies may have difficulty forecasting certain credit-negative outcomes, such as adverse litigation and leveraging corporate transactions, and such low probability/high severity events as earthquakes and hurricanes.
- The role of corporate credit analysis is to assess the company's ability to make timely payments of interest and to repay principal at maturity.
- Credit analysis is similar to equity analysis. It is important to understand, however, that bonds are contracts and that management's duty to bondholders and other creditors is limited to the terms of the contract. In contrast, management's duty to shareholders is to act in their best interest by trying to maximize the value of the company—perhaps even at the expense of bondholders at times.
- Credit analysts tend to focus more on the downside risk given the asymmetry of risk/return, whereas equity analysts focus more on upside opportunity from earnings growth, and so on.
- The "4 Cs" of credit—capacity, collateral, covenants, and character—provide a useful framework for evaluating credit risk.
- Credit analysis focuses on an issuer's ability to generate cash flow. The analysis starts with an industry assessment—structure and fundamentals—and continues with an analysis of an issuer's competitive position, management strategy, and track record.
- Credit measures are used to calculate an issuer's creditworthiness, as well as to compare its credit quality with peer companies. Key credit ratios focus on leverage and interest coverage and use such measures as EBITDA, free cash flow, funds from operations, interest expense, and balance sheet debt.

- An issuer's ability to access liquidity is also an important consideration in credit analysis.
- The higher the credit risk, the greater the offered/required yield and potential return demanded by investors. Over time, bonds with more credit risk offer higher returns but with greater volatility of return than bonds with lower credit risk.
- The yield on a credit-risky bond comprises the yield on a default risk–free bond with a comparable maturity plus a yield premium, or "spread," that comprises a credit spread and a liquidity premium. That spread is intended to compensate investors for credit risk—risk of default and loss severity in the event of default—and the credit-related risks that can cause spreads to widen and prices to decline—downgrade or credit migration risk and market liquidity risk.

  Yield spread = Liquidity premium + Credit spread.

- In times of financial market stress, the liquidity premium can increase sharply, causing spreads to widen on all credit-risky bonds, with lower-quality issuers most affected. In times of credit improvement or stability, however, credit spreads can narrow sharply as well, providing attractive investment returns.
- The impact of spread changes on holding period returns for credit-risky bonds is a product of two primary factors: the basis point spread change and the sensitivity of price to yield as reflected by (end-of-period) modified duration and convexity. Spread narrowing enhances holding period returns, whereas spread widening has a negative impact on holding period returns. Longer-duration bonds have greater price and return sensitivity to changes in spread than shorter-duration bonds.

  $$\text{Price impact} \approx -(\text{AnnModDur} \times \Delta\text{Spread}) + \tfrac{1}{2}\text{AnnConvexity} \times (\Delta\text{Spread})^2.$$

- For high-yield bonds, with their greater risk of default, more emphasis should be placed on an issuer's sources of liquidity and its debt structure and corporate structure. Credit risk can vary greatly across an issuer's debt structure depending on the seniority ranking. Many high-yield companies have complex capital structures, resulting in different levels of credit risk depending on where the debt resides.
- Covenant analysis is especially important for high-yield bonds. Key covenants include payment restrictions, limitation on liens, change of control, coverage maintenance tests (often limited to bank loans), and any guarantees from restricted subsidiaries. Covenant language can be very technical and legalistic, so it may help to seek legal or expert assistance.
- An equity-like approach to high-yield analysis can be helpful. Calculating and comparing enterprise value with EBITDA and debt/EBITDA can show a level of equity "cushion" or support beneath an issuer's debt.
- Sovereign credit analysis includes assessing both an issuer's ability and willingness to pay its debt obligations. Willingness to pay is important because, due to sovereign immunity, a sovereign government cannot be forced to pay its debts.
- In assessing sovereign credit risk, a helpful framework is to focus on five broad areas: (1) institutional effectiveness and political risks, (2) economic structure and growth prospects, (3) external liquidity and international investment position, (4) fiscal performance, flexibility, and debt burden, and (5) monetary flexibility.
- Among the characteristics of a high-quality sovereign credit are the absence of corruption and/or challenges to political framework; governmental checks and balances; respect for rule of law and property rights; commitment to honor

debts; high per capita income with stable, broad-based growth prospects; control of a reserve or actively traded currency; currency flexibility; low foreign debt and foreign financing needs relative to receipts in foreign currencies; stable or declining ratio of debt to GDP; low debt service as a percentage of revenue; low ratio of net debt to GDP; operationally independent central bank; track record of low and stable inflation; and a well-developed banking system and active money market.

- Non-sovereign or local government bonds, including municipal bonds, are typically either general obligation bonds or revenue bonds.
- General obligation (GO) bonds are backed by the taxing authority of the issuing non-sovereign government. The credit analysis of GO bonds has some similarities to sovereign analysis—debt burden per capita versus income per capita, tax burden, demographics, and economic diversity. Underfunded and "off-balance-sheet" liabilities, such as pensions for public employees and retirees, are debt-like in nature.
- Revenue-backed bonds support specific projects, such as toll roads, bridges, airports, and other infrastructure. The creditworthiness comes from the revenues generated by usage fees and tolls levied.

PRACTICE PROBLEMS

1 The risk that a bond's creditworthiness declines is *best* described by:
   - **A** credit migration risk.
   - **B** market liquidity risk.
   - **C** spread widening risk.

2 Stedsmart Ltd and Fignermo Ltd are alike with respect to financial and operating characteristics, except that Stedsmart Ltd has less publicly traded debt outstanding than Fignermo Ltd. Stedsmart Ltd is *most likely* to have:
   - **A** no market liquidity risk.
   - **B** lower market liquidity risk.
   - **C** higher market liquidity risk.

3 In the event of default, the recovery rate of which of the following bonds would *most likely* be the highest?
   - **A** First mortgage debt
   - **B** Senior unsecured debt
   - **C** Junior subordinate debt

4 During bankruptcy proceedings of a firm, the priority of claims was not strictly adhered to. Which of the following is the *least likely* explanation for this outcome?
   - **A** Senior creditors compromised.
   - **B** The value of secured assets was less than the amount of the claims.
   - **C** A judge's order resulted in actual claims not adhering to strict priority of claims.

5 A fixed-income analyst is *least likely* to conduct an independent analysis of credit risk because credit rating agencies:
   - **A** may at times mis-rate issues.
   - **B** often lag the market in pricing credit risk.
   - **C** cannot foresee future debt-financed acquisitions.

6 If goodwill makes up a large percentage of a company's total assets, this *most likely* indicates that:
   - **A** the company has low free cash flow before dividends.
   - **B** there is a low likelihood that the market price of the company's common stock is below book value.
   - **C** a large percentage of the company's assets are not of high quality.

7 In order to analyze the **collateral** of a company, a credit analyst should assess the:
   - **A** cash flows of the company.
   - **B** soundness of management's strategy.
   - **C** value of the company's assets in relation to the level of debt.

8 In order to determine the **capacity** of a company, it would be *most* appropriate to analyze the:
   - **A** company's strategy.

B growth prospects of the industry.

C aggressiveness of the company's accounting policies.

9 A credit analyst is evaluating the credit worthiness of three companies: a construction company, a travel and tourism company, and a beverage company. Both the construction and travel and tourism companies are cyclical, whereas the beverage company is non-cyclical. The construction company has the highest debt level of the three companies. The highest credit risk is *most likely* exhibited by the:

A construction company.

B beverage company.

C travel and tourism company.

10 Based on the information provided in Exhibit 1, the EBITDA interest coverage ratio of Adidas AG is *closest* to:

A 16.02x.

B 23.34x.

C 37.98x.

Exhibit 1 Adidas AG Excerpt from Consolidated Income Statement in a given year (€ in millions)

Gross profit	12,293
Royalty and commission income	154
Other operating income	56
Other operating expenses	9,843
Operating profit	2,660
Interest income	64
Interest expense	166
Income before taxes	2,558
Income taxes	640
Net income	1,918

Additional information:
Depreciation and amortization: €1,214 million

Source: Adidas AG Annual Financial Statements, December 2019.

11 The following information is from the annual report of Adidas AG for December 2019:

- Depreciation and amortization: €1,214 million
- Total assets: €20,640 million
- Total debt: €4,364 million
- Shareholders' equity: €7,058 million

The debt/capital of Adidas AG is *closest* to:

A 21.14%.

B 38.21%.

C 61.83%.

12 Funds from operations (FFO) of Pay Handle Ltd (a fictitious company) increased in 20X1. In 20X1, the total debt of the company remained unchanged while additional common shares were issued. Pay Handle Ltd's ability to service its debt in 20X1, as compared to 20X0, *most likely*:

A improved.

B worsened.

C remained the same.

13 Based on the information in Exhibit 2, GZ Group's (a hypothetical company) credit risk is *most likely*:

A lower than the industry.

B higher than the industry.

C the same as the industry.

Exhibit 2 European Food, Beverage, and Tobacco Industry and GZ Group Selected Financial Ratios for 20X0

	Total Debt/Total Capital (%)	FFO/Total Debt (%)	Return on Capital (%)	Total Debt/ EBITDA (x)	EBITDA Interest Coverage (x)
GZ Group	47.1	77.5	19.6	1.2	17.7
Industry median	**42.4**	**23.6**	**6.55**	**2.85**	**6.45**

14 Based on the information in Exhibit 3, the credit rating of DCM Group (a hypothetical company in the European food & beverage sector) is *most likely*:

A lower than AB plc.

B higher than AB plc.

C the same as AB plc.

Exhibit 3 DCM Group and AB plc Selected Financial Ratios for 20X0

Company	Total Debt/ Total Capital (%)	FFO/Total Debt (%)	Return on Capital (%)	Total Debt/EBITDA (x)	EBITDA Interest Coverage (x)
AB plc	0.2	84.3	0.1	1.0	13.9
DCM Group	42.9	22.9	8.2	3.2	3.2
European Food, Beverage, and Tobacco median	**42.4**	**23.6**	**6.55**	**2.85**	**6.45**

15 Holding all other factors constant, the *most likely* effect of low demand and heavy new issue supply on bond yield spreads is that yield spreads will:

A widen.

B tighten.

C not be affected.

16 Credit risk of a corporate bond is *best* described as the:

A risk that an issuer's creditworthiness deteriorates.

B probability that the issuer fails to make full and timely payments.

C risk of loss resulting from the issuer failing to make full and timely payments.

17 The risk that the price at which investors can actually transact differs from the quoted price in the market is called:

A spread risk.

B credit migration risk.

C market liquidity risk.

18 Loss severity is *best* described as the:

A default probability multiplied by the loss given default.

B portion of a bond's value recovered by bondholders in the event of default.

C portion of a bond's value, including unpaid interest, an investor loses in the event of default.

19 The two components of credit risk are default probability and:

A spread risk.

B loss severity.

C market liquidity risk.

20 For a high-quality debt issuer with a large amount of publicly traded debt, bond investors tend to devote *most* effort to assessing the issuer's:

A default risk.

B loss severity.

C market liquidity risk.

21 The expected loss for a given debt instrument is estimated as the product of default probability and:

A (1 + Recovery rate).

B (1 – Recovery rate).

C 1/(1 + Recovery rate).

22 The priority of claims for senior subordinated debt is:

A lower than for senior unsecured debt.

B the same as for senior unsecured debt.

C higher than for senior unsecured debt.

23 A senior unsecured credit instrument holds a higher priority of claims than one ranked as:

A mortgage debt.

B second lien loan.

C senior subordinated.

24 In a bankruptcy proceeding, when the absolute priority of claims is enforced:

A senior subordinated creditors rank above second lien holders.

B preferred equity shareholders rank above unsecured creditors.

C creditors with a secured claim have the first right to the value of that specific property.

25 In the event of default, which of the following is *most likely* to have the highest recovery rate?

A Second lien

B Senior unsecured

C Senior subordinated

26 The process of moving credit ratings of different issues up or down from the issuer rating in response to different payment priorities is *best* described as:

A notching.

B structural subordination.

C cross-default provisions.

27 The factor considered by rating agencies when a corporation has debt at both its parent holding company and operating subsidiaries is *best* referred to as:

A credit migration risk.

B corporate family rating.

C structural subordination.

28 Which type of security is *most likely* to have the same rating as the issuer?

A Preferred stock

B Senior secured bond

C Senior unsecured bond

29 Which of the following corporate debt instruments has the highest seniority ranking?

A Second lien

B Senior unsecured

C Senior subordinated

30 An issuer credit rating usually applies to a company's:

A secured debt.

B subordinated debt.

C senior unsecured debt.

31 The rating agency process whereby the credit ratings on issues are moved up or down from the issuer rating *best* describes:

A notching.

B pari passu ranking.

C cross-default provisions.

32 The notching adjustment for corporate bonds rated Aa2/AA is *most likely*:

A larger than the notching adjustment for corporate bonds rated B2/B.

B the same as the notching adjustment for corporate bonds rated B2/B.

C smaller than the notching adjustment for corporate bonds rated B2/B.

33 Which of the following statements about credit ratings is *most accurate*?

A Credit ratings can migrate over time.

B Changes in bond credit ratings precede changes in bond prices.

C Credit ratings are focused on expected loss rather than risk of default.

34 Which industry characteristic *most likely* has a positive effect on a company's ability to service debt?

A Low barriers to entry in the industry

B High number of suppliers to the industry

C Broadly dispersed market share among large number of companies in the industry

35 When determining the capacity of a borrower to service debt, a credit analyst should begin with an examination of:

A industry structure.

B industry fundamentals.

C company fundamentals.

36 Which of the following accounting issues should *mostly likely* be considered a character warning flag in credit analysis?

A Expensing items immediately

B Changing auditors infrequently

C Significant off-balance-sheet financing

37 In credit analysis, capacity is *best* described as the:

A quality of management.

B ability of the borrower to make its debt payments on time.

C quality and value of the assets supporting an issuer's indebtedness.

38 Among the four Cs of credit analysis, the recognition of revenue prematurely *most likely* reflects a company's:

A character.

B covenants.

C collateral.

Use the following Exhibit for Questions 39 and 40

Exhibit 4 Industrial Comparative Ratio Analysis, Year 20XX

	EBITDA Margin (%)	Return on Capital (%)	EBIT/ Interest Expense (×)	EBITDA/ Interest Expense (×)	Debt/ EBITDA (×)	Debt/ Capital (%)
Company A	25.1	25.0	15.9	19.6	1.6	35.2
Company B	29.6	36.3	58.2	62.4	0.5	15.9
Company C	21.8	16.6	8.9	12.4	2.5	46.3

39 Based on only the leverage ratios in Exhibit 4, the company with the *highest* credit risk is:

A Company A.

B Company B.

C Company C.

40 Based on only the coverage ratios in Exhibit 4, the company with the *highest* credit quality is:

A Company A.

B Company B.
C Company C.

Use the following Exhibits for Questions 41 and 42

Exhibit 5 Consolidated Income Statement (£ millions)

	Company X	Company Y
Net revenues	50.7	83.7
Operating expenses	49.6	70.4
Operating income	1.1	13.3
Interest income	0.0	0.0
Interest expense	0.6	0.8
Income before income taxes	0.5	12.5
Provision for income taxes	−0.2	−3.5
Net income	0.3	9.0

Exhibit 6 Consolidated Balance Sheets (£ millions)

	Company X	Company Y
ASSETS		
Current assets	10.3	21.9
Property, plant, and equipment, net	3.5	20.1
Goodwill	8.3	85.0
Other assets	0.9	5.1
Total assets	23.0	132.1
LIABILITIES AND SHAREHOLDERS' EQUITY		
Current liabilities		
Accounts payable and accrued expenses	8.4	16.2
Short-term debt	0.5	8.7
Total current liabilities	8.9	24.9
Long-term debt	11.7	21.1
Other non-current liabilities	1.1	22.1

(continued)

Exhibit 6 (Continued)

	Company X	Company Y
Total liabilities	21.7	68.1
Total shareholders' equity	1.3	64.0
Total liabilities and shareholders' equity	23.0	132.1

Exhibit 7 Consolidated Statements of Cash Flow (£ millions)

	Company X	Company Y
CASH FLOWS FROM OPERATING ACTIVITIES		
Net income	0.3	9.0
Depreciation	1.0	3.8
Goodwill impairment	2.0	1.6
Changes in working capital	0.0	–0.4
Net cash provided by operating activities	3.3	14.0
CASH FLOWS FROM INVESTING ACTIVITIES		
Additions to property and equipment	–1.0	–4.0
Additions to marketable securities	–0.1	0.0
Proceeds from sale of property and equipment	0.2	2.9
Proceeds from sale of marketable securities	0.3	0.0
Net cash used in investing activities	–0.6	–1.1
CASH FLOWS FROM FINANCING ACTIVITIES		
Repurchase of common stock	–1.5	–4.0
Dividends to shareholders	–0.3	–6.1
Change in short-term debt	0.0	–3.4
Additions to long-term debt	3.9	3.9
Reductions in long-term debt	–3.4	–2.5
Net cash–financing activities	–1.3	–12.1
NET INCREASE IN CASH AND CASH EQUIVALENTS	1.4	0.8

41 Based on Exhibits 5–7, in comparison to Company X, Company Y has a higher:

A debt/capital.

B debt/EBITDA.

C free cash flow after dividends/debt.

42 Based on Exhibits 5–7, in comparison to Company Y, Company X has greater:

A leverage.

B interest coverage.

C operating profit margin.

43 Credit yield spreads *most likely* widen in response to:

A high demand for bonds.

B weak performance of equities.

C strengthening economic conditions.

44 The factor that *most likely* results in corporate credit spreads widening is:

A an improving credit cycle.

B weakening economic conditions.

C a period of high demand for bonds.

45 Credit spreads are *most likely* to widen:

A in a strengthening economy.

B as the credit cycle improves.

C in periods of heavy new issue supply and low borrower demand.

46 Which of the following factors in credit analysis is more important for general obligation non-sovereign government debt than for sovereign debt?

A Per capita income

B Power to levy and collect taxes

C Requirement to balance an operating budget

47 In contrast to high-yield credit analysis, investment-grade analysis is *more likely* to rely on:

A spread risk.

B an assessment of bank credit facilities.

C matching of liquidity sources to upcoming debt maturities.

48 Which of the following factors would *best* justify a decision to avoid investing in a country's sovereign debt?

A Freely floating currency

B A population that is not growing

C Suitable checks and balances in policymaking

SOLUTIONS

1 A is correct. Credit migration risk or downgrade risk refers to the risk that a bond issuer's creditworthiness may deteriorate or migrate lower. The result is that investors view the risk of default to be higher, causing the spread on the issuer's bonds to widen.

2 C is correct. Market liquidity risk refers to the risk that the price at which investors transact may be different from the price indicated in the market. Market liquidity risk is increased by (1) less debt outstanding and/or (2) a lower issue credit rating. Because Stedsmart Ltd is comparable to Fignermo Ltd except for less publicly traded debt outstanding, it should have higher market liquidity risk.

3 A is correct. First mortgage debt is senior secured debt and has the highest priority of claims. First mortgage debt also has the highest expected recovery rate. First mortgage debt refers to the pledge of specific property. Neither senior unsecured nor junior subordinate debt has any claims on specific assets.

4 B is correct. Whether or not secured assets are sufficient for the claims against them does not influence priority of claims. Any deficiency between pledged assets and the claims against them becomes senior unsecured debt and still adheres to the guidelines of priority of claims.

5 C is correct. Both analysts and rating agencies have difficulty foreseeing future debt-financed acquisitions.

6 C is correct. Goodwill is viewed as a lower quality asset compared with tangible assets that can be sold and more easily converted into cash.

7 C is correct. The value of assets in relation to the level of debt is important to assess the collateral of the company—that is, the quality and value of the assets that support the debt levels of the company.

8 B is correct. The growth prospects of the industry provide the analyst insight regarding the capacity of the company.

9 A is correct. The construction company is both highly leveraged, which increases credit risk, and in a highly cyclical industry, which results in more volatile earnings.

10 B is correct. The interest expense is €166 million and EBITDA = Operating profit + Depreciation and amortization = €2,660 + 1,214 million = €3,874 million. EBITDA interest coverage = EBITDA/Interest expense = 3,874/166 = 23.34 times.

11 B is correct. Total debt is €4,364 million with Total capital = Total debt + Shareholders' equity = €4,364 + 7,058 = €11,422 million. The Debt/Capital = 4,364/11,422 = 38.21%.

12 A is correct. If the debt of the company remained unchanged but FFO increased, more cash was available to service debt compared to the previous year. Additionally, debt/capital improved, which implies that the ability of Pay Handle Ltd to service their debt also improved.

13 A is correct. Based on four of the five credit ratios, GZ Group's credit quality is superior to that of the industry.

14 A is correct. DCM Group has more financial leverage and less interest coverage than AB plc, which implies greater credit risk.

15 A is correct. Low demand implies wider yield spreads, and heavy supply widens spreads even further.

16 C is correct. Credit risk is the risk of loss resulting from the borrower failing to make full and timely payments of interest and/or principal.

17 C is correct. Market liquidity risk is the risk that the price at which investors can actually transact—buying or selling—may differ from the price indicated in the market.

18 C is correct. Loss severity is the portion of a bond's value (including unpaid interest) an investor loses in the event of default.

19 B is correct. The two components of credit risk are default probability and loss severity. In the event of default, loss severity is the portion of a bond's value (including unpaid interest) an investor loses. A and C are incorrect because spread and market liquidity risk are credit-related risks, not components of credit risk.

20 A is correct. Credit risk has two components: default risk and loss severity. Because default risk is quite low for most high-quality debt issuers, bond investors tend to focus more on this likelihood and less on the potential loss severity.

21 B is correct. The expected loss for a given debt instrument is the default probability multiplied by the loss severity given default. The loss severity is often expressed as (1 – Recovery rate).

22 A is correct. Senior subordinated debt is ranked lower than senior unsecured debt and thus has a lower priority of payment.

23 C is correct. The highest-ranked unsecured debt is senior unsecured debt. Lower-ranked debt includes senior subordinated debt. A and B are incorrect because mortgage debt and second lien loans are secured and higher ranked.

24 C is correct. According to the absolute priority of claims, in the event of bankruptcy, creditors with a secured claim have the right to the value of that specific property before any other claim.

25 A is correct. A second lien has a secured interest in the pledged assets. Second lien debt ranks higher in priority of payment than senior unsecured and senior subordinated debt and thus would most likely have a higher recovery rate.

26 A is correct. Notching is the process for moving ratings up or down relative to the issuer rating when rating agencies consider secondary factors, such as priority of claims in the event of a default and the potential loss severity.

27 C is correct. Structural subordination can arise when a corporation with a holding company structure has debt at both its parent holding company and operating subsidiaries. Debt at the operating subsidiaries is serviced by the cash flow and assets of the subsidiaries before funds are passed to the parent holding company.

28 C is correct. The issuer credit rating usually applies to its senior unsecured debt.

29 A is correct. Second lien debt is secured debt, which is senior to unsecured debt and to subordinated debt.

30 C is correct. An issuer credit rating usually applies to its senior unsecured debt.

31 A is correct. Recognizing different payment priorities, and thus the potential for higher (or lower) loss severity in the event of default, the rating agencies have adopted a notching process whereby their credit ratings on issues can be moved up or down from the issuer rating (senior unsecured).

32 C is correct. As a general rule, the higher the senior unsecured rating, the smaller the notching adjustment. Thus, for corporate bonds rated Aa2/AA, the rating agencies will typically apply smaller rating adjustments, or notches, to the related issue.

33 A is correct. Credit migration is the risk that a bond issuer's creditworthiness deteriorates, or migrates lower. Over time, credit ratings can migrate significantly from what they were at the time a bond was issued. An investor should not assume that an issuer's credit rating will remain the same from the time of purchase through the entire holding period.

34 B is correct. An industry with a high number of suppliers reduces the suppliers' negotiating power, thus helping companies control expenses and aiding in the servicing of debt.

35 A is correct. Credit analysis starts with industry structure—for example, by looking at the major forces of competition, followed by an analysis of industry fundamentals—and then turns to examination of the specific issuer.

36 C is correct. Credit analysts can make judgments about management's character by evaluating the use of aggressive accounting policies, such as timing revenue recognition. This activity is a potential warning flag for other behaviors or actions that may adversely affect an issuer's creditworthiness.

37 B is correct. Capacity refers to the ability of a borrower to service its debt. Capacity is determined through credit analysis of an issuer's industry and of the specific issuer.

38 A is correct. Credit analysts can make judgments about management's character in a number of ways, including by observing its use of aggressive accounting policies and/or tax strategies. An example of this aggressiveness is recognizing revenue prematurely.

39 C is correct. Debt/capital and debt/EBITDA are used to assess a company's leverage. Higher leverage ratios indicate more leverage and thus higher credit risk. Company C's debt/capital (46.3%) and debt/EBITDA (2.5×) are higher than those for Companies A and B.

40 B is correct. The EBITDA/interest expense and EBIT/interest expense are coverage ratios. Coverage ratios measure an issuer's ability to meet its interest payments. A higher ratio indicates better credit quality. Company B's EBITDA/interest expense (62.4×) and EBIT/interest expense (58.2×) are higher than those for Companies A and C.

41 C is correct because Company Y has a higher ratio of free cash flow after dividends to debt than Company X, not lower, as shown in the following table.

$$\text{Free cash flow after dividends as a \% of debt} = \frac{\text{FCF after dividends}}{\text{Debt}}$$

	Company X	Company Y
Cash flow from operations	£3.3	£14.0
Less		
Net capital expenditures	−0.8	−1.1
Dividends	−0.3	−6.1
Free cash flow after dividends	£2.2	£6.8
Debt	£12.2	£29.8
Free cash flow after dividends as a % of debt	(2.2/12.2) × 100	(6.8/29.8) × 100
Free cash flow after dividends as a % of debt	18.0%	22.8%

A is incorrect. Company Y has a lower debt/capital than Company X, as shown in the following table.

$$\text{Debt divided by Capital (\%)} = \frac{\text{Debt}}{(\text{Debt} + \text{Equity})}$$

	Company X	Company Y
Debt	£12.2	£29.8
Capital		
Debt	12.2	29.8
+ Equity	1.3	64.0
Capital	£13.5	£93.8
Debt/Capital (%)	(12.2/13.5) × 100	(29.8/93.8) × 100
Debt/Capital (%)	90.4%	31.8%

B is incorrect because Company Y has a lower debt/EBITDA than Company Y, not higher, as shown in the following table.

	Company X	Company Y
Operating income	£1.1	£13.3
EBIT	£1.1	£13.3
plus		
Depreciation	1.0	3.8
Amortization	0.0	0.0
EBITDA	£2.1	£17.1
Debt	£12.2	£29.8
Debt/EBITDA	12.2/2.1	29.8/17.1
Debt/EBITDA	5.81	1.74

42 A is correct. Compared with Company Y, based on both their debt/capital and their ratios of free cash flow after dividends to debt, which are measures of leverage commonly used in credit analysis, Company X is more highly leveraged, as shown in the following table.

$$\text{Debt divided by Capital (\%)} = \frac{\text{Debt}}{(\text{Debt} + \text{Equity})}$$

	Company X	Company Y
Debt	£2.2	£29.8
Capital		
Debt	2.2	29.8
+ Equity	4.3	64.0
Capital	£6.5	£93.8

(continued)

Debt/Capital (%)	(12.2/13.5) × 100	(29.8/93.8) × 100
Debt/Capital (%)	90.4%	31.8%

$$\text{Free cash flow after dividends as a \% of debt} = \frac{\text{FCF after dividends}}{\text{Debt}}$$

	Company X	Company Y
Cash flow from operations	£3.3	£14.0
Less		
Net capital expenditures	−0.8	−1.1
Dividends	−0.3	−6.1
Free cash flow after dividends	£2.2	£6.8
Debt	£12.2	£29.8
Free cash flow after dividends as a % of debt	(2.2/12.2) × 100	(6.8/29.8) × 100
Free cash flow after dividends as a % of debt	18.0%	22.8%

43 B is correct. In weak financial markets, including weak markets for equities, credit spreads will widen.

44 B is correct. Weakening economic conditions will push investors to desire a greater risk premium and drive overall credit spreads wider.

45 C is correct. In periods of heavy new issue supply, credit spreads will widen if demand is insufficient.

46 C is correct. Non-sovereign governments typically must balance their operating budgets and lack the discretion to use monetary policy as many sovereigns can.

47 A is correct. Most investors in investment-grade debt focus on spread risk—that is, the effect of changes in spreads on prices and returns—while in high-yield analysis, the focus on default risk is relatively greater.

48 B is correct. Among the most important considerations in sovereign credit analysis is growth and age distribution of population. A relatively young and growing population contributes to growth in GDP and an expanding tax base and relies less on social services, pensions, and health care relative to an older population.

Derivatives

STUDY SESSION

Study Session 15	Derivatives

TOPIC LEVEL LEARNING OUTCOME

The candidate should be able to demonstrate a working knowledge of the analysis of derivatives, including forwards, futures, options, and swaps.

Derivatives—financial instruments whose prices are derived from the value of some underlying asset—have become increasingly important for managing financial risk, exploiting investment opportunities, and creating synthetic asset class exposure. As in other security markets, arbitrage and market efficiency play a critical role in establishing prices for these securities.

DERIVATIVES

STUDY SESSION 15

Derivatives

This study session builds the conceptual framework for understanding the basic derivatives and derivative markets. Essential features and valuation concepts for forward commitments such as forwards, futures, and swaps and contingent claims such as options are introduced.

READING ASSIGNMENTS

READING

45

Derivative Markets and Instruments

by Don M. Chance, PhD, CFA

Don M. Chance, PhD, CFA, is at Louisiana State University (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** define a derivative and distinguish between exchange-traded and over-the-counter derivatives;
☐	**b.** contrast forward commitments with contingent claims;
☐	**c.** define forward contracts, futures contracts, options (calls and puts), swaps, and credit derivatives and compare their basic characteristics;
☐	**d.** determine the value at expiration and profit from a long or a short position in a call or put option;
☐	**e.** describe purposes of, and controversies related to, derivative markets;
☐	**f.** explain arbitrage and the role it plays in determining prices and promoting market efficiency.

1 DERIVATIVES: INTRODUCTION, DEFINITIONS, AND USES

a define a derivative and distinguish between exchange-traded and over-the-counter derivatives;

b contrast forward commitments with contingent claims;

Equity, fixed-income, currency, and commodity markets are facilities for trading the basic assets of an economy. Equity and fixed-income securities are claims on the assets of a company. Currencies are the monetary units issued by a government or central bank. Commodities are natural resources, such as oil or gold. These underlying assets are said to trade in **cash markets** or **spot markets** and their prices are sometimes referred to as **cash prices** or **spot prices**, though we usually just refer to them as stock prices, bond prices, exchange rates, and commodity prices. These markets exist around the world and receive much attention in the financial and mainstream media. Hence, they are relatively familiar not only to financial experts but also to the general population.

Somewhat less familiar are the markets for **derivatives**, which are financial instruments that derive their values from the performance of these basic assets. This reading is an overview of derivatives. Subsequent readings will explore many aspects of derivatives and their uses in depth. Among the questions that this first reading will address are the following:

- What are the defining characteristics of derivatives?
- What purposes do derivatives serve for financial market participants?
- What is the distinction between a forward commitment and a contingent claim?
- What are forward and futures contracts? In what ways are they alike and in what ways are they different?
- What are swaps?
- What are call and put options and how do they differ from forwards, futures, and swaps?
- What are credit derivatives and what are the various types of credit derivatives?
- What are the benefits of derivatives?
- What are some criticisms of derivatives and to what extent are they well founded?
- What is arbitrage and what role does it play in a well-functioning financial market?

This reading is organized as follows. Section 1 explores the definition and uses of derivatives and establishes some basic terminology. Section 2 describes derivatives markets. Sections 3–9 categorize and explain types of derivatives. Sections 10 and 11 discuss the benefits and criticisms of derivatives, respectively. Section 12 introduces the basic principles of derivative pricing and the concept of arbitrage.

Derivatives: Definitions and Uses

The most common definition of a derivative reads approximately as follows:

> *A derivative is a financial instrument that derives its performance from the performance of an underlying asset.*

This definition, despite being so widely quoted, can nonetheless be a bit troublesome. For example, it can also describe mutual funds and exchange-traded funds, which would never be viewed as derivatives even though they derive their values from the values of the underlying securities they hold. Perhaps the distinction that best characterizes derivatives is that they usually *transform* the performance of the underlying asset before paying it out in the derivatives transaction. In contrast, with the exception of expense deductions, mutual funds and exchange-traded funds simply pass through the returns of their underlying securities. This transformation of performance is typically understood or implicit in references to derivatives but rarely makes its way into the formal definition. In keeping with customary industry practice, this characteristic will be retained as an implied, albeit critical, factor distinguishing derivatives from mutual funds and exchange-traded funds and some other straight pass-through instruments. Also, note that the idea that derivatives take their *performance* from an underlying asset encompasses the fact that derivatives take their value and certain other characteristics from the underlying asset. Derivatives strategies perform in ways that are derived from the underlying and the specific features of derivatives.

Derivatives are similar to insurance in that both allow for the transfer of risk from one party to another. As everyone knows, insurance is a financial contract that provides protection against loss. The party bearing the risk purchases an insurance policy, which transfers the risk to the other party, the insurer, for a specified period of time. The risk itself does not change, but the party bearing it does. Derivatives allow for this same type of transfer of risk. One type of derivative in particular, the put option, when combined with a position exposed to the risk, functions almost exactly like insurance, but all derivatives can be used to protect against loss. Of course, an insurance contract must specify the underlying risk, such as property, health, or life. Likewise, so do derivatives. As noted earlier, derivatives are associated with an underlying asset. As such, the so-called "underlying asset" is often simply referred to as the **underlying**, whose value is the source of risk. In fact, the underlying need not even be an asset itself. Although common derivatives underlyings are equities, fixed-income securities, currencies, and commodities, other derivatives underlyings include interest rates, credit, energy, weather, and even other derivatives, all of which are not generally thought of as assets. Thus, like insurance, derivatives pay off on the basis of a source of risk, which is often, but not always, the value of an underlying asset. And like insurance, derivatives have a definite life span and expire on a specified date.

Derivatives are created in the form of legal contracts. They involve two parties—the buyer and the seller (sometimes known as the writer)—each of whom agrees to do something for the other, either now or later. The buyer, who purchases the derivative, is referred to as the **long** or the holder because he owns (holds) the derivative and holds a long position. The seller is referred to as the **short** because he holds a short position.[1]

A derivative contract always defines the rights and obligations of each party. These contracts are intended to be, and almost always are, recognized by the legal system as commercial contracts that each party expects to be upheld and supported in the legal system. Nonetheless, disputes sometimes arise, and lawyers, judges, and juries may be required to step in and resolve the matter.

There are two general classes of derivatives. Some provide the ability to lock in a price at which one might buy or sell the underlying. Because they force the two parties to transact in the future at a previously agreed-on price, these instruments are called **forward commitments**. The various types of forward commitments are called forward contracts, futures contracts, and swaps. Another class of derivatives provides *the right but not the obligation* to buy or sell the underlying at a pre-determined price. Because the choice of buying or selling versus doing nothing depends on a particular random outcome, these derivatives are called **contingent claims**. The primary contingent claim is called an **option**. The types of derivatives will be covered in more detail later in this reading and in considerably more depth later in the curriculum.

The existence of derivatives begs the obvious question of what purpose they serve. If one can participate in the success of a company by holding its equity, what reason can possibly explain why another instrument is required that takes its value from the performance of the equity? Although equity and other fundamental markets exist and usually perform reasonably well without derivative markets, it is possible that derivative markets can *improve* the performance of the markets for the underlyings. As you will see later in this reading, that is indeed true in practice.

Derivative markets create beneficial opportunities that do not exist in their absence. Derivatives can be used to create strategies that cannot be implemented with the underlyings alone. For example, derivatives make it easier to go short, thereby benefiting from a decline in the value of the underlying. In addition, derivatives, in and

1 In the financial world, the *long* always benefits from an increase in the value of the instrument he owns, and the *short* always benefits from a decrease in the value of the instrument he has sold. Think of the long as having possession of something and the short as having incurred an obligation to deliver that something.

of themselves, are characterized by a relatively high degree of leverage, meaning that participants in derivatives transactions usually have to invest only a small amount of their own capital relative to the value of the underlying. As such, small movements in the underlying can lead to fairly large movements in the amount of money made or lost on the derivative. Derivatives generally trade at lower transaction costs than comparable spot market transactions, are often more liquid than their underlyings, and offer a simple, effective, and low-cost way to transfer risk. For example, a shareholder of a company can reduce or even completely eliminate the market exposure by trading a derivative on the equity. Holders of fixed-income securities can use derivatives to reduce or completely eliminate interest rate risk, allowing them to focus on the credit risk. Alternatively, holders of fixed-income securities can reduce or eliminate the credit risk, focusing more on the interest rate risk. Derivatives permit such adjustments easily and quickly. These features of derivatives are covered in more detail later in this reading.

The types of performance transformations facilitated by derivatives allow market participants to practice more effective risk management. Indeed, the entire field of derivatives, which at one time was focused mostly on the instruments themselves, is now more concerned with the *uses* of the instruments. Just as a carpenter uses a hammer, nails, screws, a screwdriver, and a saw to build something useful or beautiful, a financial expert uses derivatives to manage risk. And just as it is critically important that a carpenter understand how to use these tools, an investment practitioner must understand how to properly use derivatives. In the case of the carpenter, the result is building something useful; in the case of the financial expert, the result is managing financial risk. Thus, like tools, derivatives serve a valuable purpose but like tools, they must be used carefully.

The practice of risk management has taken a prominent role in financial markets. Indeed, whenever companies announce large losses from trading, lending, or operations, stories abound about how poorly these companies managed risk. Such stories are great attention grabbers and a real boon for the media, but they often miss the point that risk management does not guarantee that large losses will not occur. Rather, **risk management** *is the process by which an organization or individual defines the level of risk it wishes to take, measures the level of risk it is taking, and adjusts the latter to equal the former*. Risk management never offers a guarantee that large losses will not occur, and it does not eliminate the possibility of total failure. To do so would typically require that the amount of risk taken be so small that the organization would be effectively constrained from pursuing its primary objectives. Risk taking is inherent in all forms of economic activity and life in general. The possibility of failure is never eliminated.

EXAMPLE 1

Characteristics of Derivatives

1 Which of the following is the best example of a derivative?
   - A A global equity mutual fund
   - B A non-callable government bond
   - C A contract to purchase Apple Computer at a fixed price

2 Which of the following is **not** a characteristic of a derivative?
   - A An underlying
   - B A low degree of leverage
   - C Two parties—a buyer and a seller

3 Which of the following statements about derivatives is **not** true?
   - A They are created in the spot market.
   - B They are used in the practice of risk management.
   - C They take their values from the value of something else.

Solution to 1:

C is correct. Mutual funds and government bonds are not derivatives. A government bond is a fundamental asset on which derivatives might be created, but it is not a derivative itself. A mutual fund can technically meet the definition of a derivative, but as noted in the reading, derivatives transform the value of a payoff of an underlying asset. Mutual funds merely pass those payoffs through to their holders.

Solution to 2:

B is correct. All derivatives have an underlying and must have a buyer and a seller. More importantly, derivatives have high degrees of leverage, not low degrees of leverage.

Solution to 3:

A is correct. Derivatives are used to practice risk management and they take (derive) their values from the value of something else, the underlying. They are not created in the spot market, which is where the underlying trades.

Note also that risk management is a dynamic and ongoing process, reflecting the fact that the risk assumed can be difficult to measure and is constantly changing. As noted, derivatives are tools, indeed *the* tools that make it easier to manage risk. Although one can trade stocks and bonds (the underlyings) to adjust the level of risk, it is almost always more effective to trade derivatives.

Risk management is addressed more directly elsewhere in the CFA curriculum, but the study of derivatives necessarily entails the concept of risk management. In an explanation of derivatives, the focus is usually on the instruments and it is easy to forget the overriding objective of managing risk. Unfortunately, that would be like a carpenter obsessed with his hammer and nails, forgetting that he is building a piece of furniture. It is important to always try to keep an eye on the objective of managing risk.

2 THE STRUCTURE OF DERIVATIVE MARKETS

- a define a derivative and distinguish between exchange-traded and over-the-counter derivatives;
- b contrast forward commitments with contingent claims;

Having an understanding of equity, fixed-income, and currency markets is extremely beneficial—indeed, quite necessary—in understanding derivatives. One could hardly consider the wisdom of using derivatives on a share of stock if one did not understand the equity markets reasonably well. As you likely know, equities trade on organized exchanges as well as in over-the-counter (OTC) markets. These exchange-traded equity markets—such as the Deutsche Börse, the Tokyo Stock Exchange, and the New York Stock Exchange and its Eurex affiliate—are formal organizational structures that bring buyers and sellers together through market makers, or dealers, to facilitate transactions. Exchanges have formal rule structures and are required to comply with all securities laws.

OTC securities markets operate in much the same manner, with similar rules, regulations, and organizational structures. At one time, the major difference between OTC and exchange markets for securities was that the latter brought buyers and sellers together in a physical location, whereas the former facilitated trading strictly in an electronic manner. Today, these distinctions are blurred because many organized securities exchanges have gone completely to electronic systems. Moreover, OTC securities markets can be formally organized structures, such as NASDAQ, or can merely refer to informal networks of parties who buy and sell with each other, such as the corporate and government bond markets in the United States.

The derivatives world also comprises organized exchanges and OTC markets. Although the derivatives world is also moving toward less distinction between these markets, there are clear differences that are important to understand.

2.1 Exchange-Traded Derivatives Markets

Derivative instruments are created and traded either on an exchange or on the OTC market. Exchange-traded derivatives are standardized, whereas OTC derivatives are customized. To standardize a derivative contract means that its terms and conditions are precisely specified by the exchange and there is very limited ability to alter those terms. For example, an exchange might offer trading in certain types of derivatives that expire only on the third Friday of March, June, September, and December. If a party wanted the derivative to expire on any other day, it would not be able to trade such a derivative on that exchange, nor would it be able to persuade the exchange to create it, at least not in the short run. If a party wanted a derivative on a particular entity, such as a specific stock, that party could trade it on that exchange only if the exchange had specified that such a derivative could trade. Even the magnitudes of the contracts are specified. If a party wanted a derivative to cover €150,000 and the exchange specified that contracts could trade only in increments of €100,000, the party could do nothing about it if it wanted to trade that derivative on that exchange.

This standardization of contract terms facilitates the creation of a more liquid market for derivatives. If all market participants know that derivatives on the euro trade in 100,000-unit lots and that they all expire only on certain days, the market functions more effectively than it would if there were derivatives with many different unit sizes and expiration days competing in the same market at the same time. This standardization makes it easier to provide liquidity. Through designated market makers, derivatives exchanges guarantee that derivatives can be bought and sold.[2]

The cornerstones of the exchange-traded derivatives market are the market makers (or dealers) and the speculators, both of whom typically own memberships on the exchange.[3] The market makers stand ready to buy at one price and sell at a higher price. With standardization of terms and an active market, market makers are often able to buy and sell almost simultaneously at different prices, locking in small,

2 It is important to understand that merely being able to buy and sell a derivative, or even a security, does not mean that liquidity is high and that the cost of liquidity is low. Derivatives exchanges guarantee that a derivative can be bought and sold, but they do not guarantee the price. The ask price (the price at which the market maker will sell) and the bid price (the price at which the market maker will buy) can be far apart, which they will be in a market with low liquidity. Hence, such a market can have liquidity, loosely defined, but the cost of liquidity can be quite high. The factors that can lead to low liquidity for derivatives are similar to those for securities: little trading interest and a high level of uncertainty.

3 Exchanges are owned by their *members*, whose memberships convey the right to trade. In addition, some exchanges are themselves publicly traded corporations whose members are shareholders, and there are also non-member shareholders.

short-term profits—a process commonly known as scalping. In some cases, however, they are unable to do so, thereby forcing them to either hold exposed positions or find other parties with whom they can trade and thus lay off (get rid of) the risk. This is when speculators come in. Although speculators are market participants who are willing to take risks, it is important to understand that being a speculator does not mean the reckless assumption of risk. Although speculators will take large losses at times, good speculators manage those risks by watching their exposures, absorbing market information, and observing the flow of orders in such a manner that they are able to survive and profit. Often, speculators will hedge their risks when they become uncomfortable.

Standardization also facilitates the creation of a clearing and settlement operation. **Clearing** refers to the process by which the exchange verifies the execution of a transaction and records the participants' identities. **Settlement** refers to the related process in which the exchange transfers money from one participant to the other or from a participant to the exchange or vice versa. This flow of money is a critical element of derivatives trading. Clearly, there would be no confidence in markets in which money is not efficiently collected and disbursed. Derivatives exchanges have done an excellent job of clearing and settlement, especially in comparison to securities exchanges. Derivatives exchanges clear and settle all contracts overnight, whereas most securities exchanges require two business days.

The clearing and settlement process of derivative transactions also provides a credit guarantee. If two parties engage in a derivative contract on an exchange, one party will ultimately make money and the other will lose money. Derivatives exchanges use their clearinghouses to provide a guarantee to the winning party that if the loser does not pay, the clearinghouse will pay the winning party. The clearinghouse is able to provide this credit guarantee by requiring a cash deposit, usually called the **margin bond** or **performance bond**, from the participants to the contract. Derivatives clearinghouses manage these deposits, occasionally requiring additional deposits, so effectively that they have never failed to pay in the nearly 100 years they have existed. We will say more about this process later and illustrate how it works.

Exchange markets are said to have **transparency**, which means that full information on all transactions is disclosed to exchanges and regulatory bodies. All transactions are centrally reported within the exchanges and their clearinghouses, and specific laws require that these markets be overseen by national regulators. Although this would seem a strong feature of exchange markets, there is a definite cost. Transparency means a loss of privacy: National regulators can see what transactions have been done. Standardization means a loss of flexibility: A participant can do only the transactions that are permitted on the exchange. Regulation means a loss of both privacy and flexibility. It is not that transparency or regulation is good and the other is bad. It is simply a trade-off.

Derivatives exchanges exist in virtually every developed (and some emerging market) countries around the world. Some exchanges specialize in derivatives and others are integrated with securities exchanges.

Although there have been attempts to create somewhat non-standardized derivatives for trading on an exchange, such attempts have not been particularly successful. Standardization is a critical element by which derivatives exchanges are able to provide their services. We will look at this point again when discussing the alternative to standardization: customized OTC derivatives.

2.2 Over-the-Counter Derivatives Markets

The OTC derivatives markets comprise an informal network of market participants that are willing to create and trade virtually any type of derivative that can legally exist. The backbone of these markets is the set of dealers, which are typically banks. Most of these banks are members of a group called the International Swaps and Derivatives Association (ISDA), a worldwide organization of financial institutions that engage in derivative transactions, primarily as dealers. As such, these markets are sometimes called *dealer markets*. Acting as principals, these dealers informally agree to buy and sell various derivatives. It is *informal* because the dealers are not obligated to do so. Their participation is based on a desire to profit, which they do by purchasing at one price and selling at a higher price. Although it might seem that a dealer who can "buy low, sell high" could make money easily, the process in practice is not that simple. Because OTC instruments are not standardized, a dealer cannot expect to buy a derivative at one price and simultaneously sell it to a different party who happens to want to buy the same derivative at the same time and at a higher price.

To manage the risk they assume by buying and selling customized derivatives, OTC derivatives dealers typically hedge their risks by engaging in alternative but similar transactions that pass the risk on to other parties. For example, if a company comes to a dealer to buy a derivative on the euro, the company would effectively be transferring the risk of the euro to the dealer. The dealer would then attempt to lay off (get rid of) that risk by engaging in an alternative but similar transaction that would transfer the risk to another party. This hedge might involve another derivative on the euro or it might simply be a transaction in the euro itself. Of course, that begs the question of why the company could not have laid off the risk itself and avoided the dealer. Indeed, some can and do, but laying off risk is not simple. Unable to find identical offsetting transactions, dealers usually have to find *similar* transactions with which they can lay off the risk. Hedging one derivative with a different kind of derivative on the same underlying is a similar but not identical transaction. It takes specialized knowledge and complex models to be able to do such transactions effectively, and dealers are more capable of doing so than are ordinary companies. Thus, one might think of a dealer as a middleman, a sort of financial wholesaler using its specialized knowledge and resources to facilitate the transfer of risk. In the same manner that one could theoretically purchase a consumer product from a manufacturer, a network of specialized middlemen and retailers is often a more effective method.

Because of the customization of OTC derivatives, there is a tendency to think that the OTC market is less liquid than the exchange market. That is not necessarily true. Many OTC instruments can easily be created and then essentially offset by doing the exact opposite transaction, often with the same party. For example, suppose Corporation A buys an OTC derivative from Dealer B. Before the expiration date, Corporation A wants to terminate the position. It can return to Dealer B and ask to sell a derivative with identical terms. Market conditions will have changed, of course, and the value of the derivative will not be the same, but the transaction can be conducted quite easily with either Corporation A or Dealer B netting a gain at the expense of the other. Alternatively, Corporation A could do this transaction with a different dealer, the result of which would remove exposure to the underlying risk but would leave two transactions open and some risk that one party would default to the other. In contrast to this type of OTC liquidity, some exchange-traded derivatives have very little trading interest and thus relatively low liquidity. Liquidity is always driven by trading interest, which can be strong or weak in both types of markets.

OTC derivative markets operate at a lower degree of regulation and oversight than do exchange-traded derivative markets. In fact, until around 2010, it could largely be said that the OTC market was essentially unregulated. OTC transactions could be executed with only the minimal oversight provided through laws that regulated the parties themselves, not the specific instruments. Following the financial crisis of 2007–2009, new regulations began to blur the distinction between OTC and exchange-listed markets. In both the United States (the Wall Street Reform and Consumer Protection Act of 2010, commonly known as the Dodd–Frank Act) and Europe (the Regulation of the European Parliament and of the Council on OTC Derivatives, Central

Counterparties, and Trade Repositories), regulations are changing the characteristics of OTC markets. In general, world policy-makers have advanced an agenda to make global derivatives markets more resilient and robust, pursuing increased transparency and lowered systemic risk.

When the full implementation of these new laws takes place, a number of OTC transactions will have to be cleared through central clearing agencies, information on most OTC transactions will need to be reported to regulators, and entities that operate in the OTC market will be more closely monitored. There are, however, quite a few exemptions that cover a significant percentage of derivative transactions. Clearly, the degree of OTC regulation, although increasing in recent years, is still lighter than that of exchange-listed market regulation. Many transactions in OTC markets will retain a degree of privacy with lower transparency, and most importantly, the OTC markets will remain considerably more flexible than the exchange-listed markets.

EXAMPLE 2

Exchange-Traded versus Over-the-Counter Derivatives

1. Which of the following characteristics is **not** associated with exchange-traded derivatives?
   - **A** Margin or performance bonds are required.
   - **B** The exchange guarantees all payments in the event of default.
   - **C** All terms except the price are customized to the parties' individual needs.
2. Which of the following characteristics is associated with over-the-counter derivatives?
   - **A** Trading occurs in a central location.
   - **B** They are more regulated than exchange-listed derivatives.
   - **C** They are less transparent than exchange-listed derivatives.
3. Market makers earn a profit in both exchange and over-the-counter derivatives markets by:
   - **A** charging a commission on each trade.
   - **B** a combination of commissions and markups.
   - **C** buying at one price, selling at a higher price, and hedging any risk.
4. Which of the following statements *most* accurately describes exchange-traded derivatives relative to over-the-counter derivatives? Exchange-traded derivatives are more likely to have:
   - **A** greater credit risk.
   - **B** standardized contract terms.
   - **C** greater risk management uses.

Solution to 1:

C is correct. Exchange-traded contracts are standardized, meaning that the exchange determines the terms of the contract except the price. The exchange guarantees against default and requires margins or performance bonds.

Solution to 2:

C is correct. OTC derivatives have a lower degree of transparency than exchange-listed derivatives. Trading does not occur in a central location but, rather, is quite dispersed. Although new national securities laws are tightening the regulation of OTC derivatives, the degree of regulation is less than that of exchange-listed derivatives.

Solution to 3:

C is correct. Market makers buy at one price (the bid), sell at a higher price (the ask), and hedge whatever risk they otherwise assume. Market makers do not charge a commission. Hence, A and B are both incorrect.

Solution to 4:

B is correct. Standardization of contract terms is a characteristic of exchange-traded derivatives. A is incorrect because credit risk is well-controlled in exchange markets. C is incorrect because the risk management uses are not limited by being traded over the counter.

3 TYPES OF DERIVATIVES: INTRODUCTION, FORWARD CONTRACTS

b contrast forward commitments with contingent claims;

c define forward contracts, futures contracts, options (calls and puts), swaps, and credit derivatives and compare their basic characteristics;

As previously stated, derivatives fall into two general classifications: forward commitments and contingent claims. The factor that distinguishes forward commitments from contingent claims is that forward commitments *obligate* the parties to engage in a transaction at a future date on terms agreed upon in advance, whereas contingent claims provide one party the *right but not the obligation* to engage in a future transaction on terms agreed upon in advance.

3.1 Forward Commitments

Forward commitments are contracts entered into at one point in time that require both parties to engage in a transaction at a later point in time (the expiration) on terms agreed upon at the start. The parties establish the identity and quantity of the underlying, the manner in which the contract will be executed or settled when it expires, and the fixed price at which the underlying will be exchanged. This fixed price is called the **forward price**.

As a hypothetical example of a forward contract, suppose that today Markus and Johannes enter into an agreement that Markus will sell his BMW to Johannes for a price of €30,000. The transaction will take place on a specified date, say, 180 days from today. At that time, Markus will deliver the vehicle to Johannes's home and Johannes will give Markus a bank-certified check for €30,000. There will be no recourse, so if the vehicle has problems later, Johannes cannot go back to Markus for compensation. It should be clear that both Markus and Johannes must do their due diligence and carefully consider the reliability of each other. The car could have serious quality issues and Johannes could have financial problems and be unable to pay the €30,000. Obviously, the transaction is essentially unregulated. Either party could renege on his obligation, in response to which the other party could go to court, provided a formal contract exists and is carefully written. Note finally that one of the two parties is likely to end up gaining and the other losing, depending on the secondary market price of this type of vehicle at expiration of the contract.

This example is quite simple but illustrates the essential elements of a forward contract. In the financial world, such contracts are very carefully written, with legal provisions that guard against fraud and require extensive credit checks. Now let us take a deeper look at the characteristics of forward contracts.

3.1.1 *Forward Contracts*

The following is the formal definition of a forward contract:

> *A forward contract is an over-the-counter derivative contract in which two parties agree that one party, the buyer, will purchase an underlying asset from the other party, the seller, at a later date at a fixed price they agree on when the contract is signed.*

In addition to agreeing on the price at which the underlying asset will be sold at a later date, the two parties also agree on several other matters, such as the specific identity of the underlying, the number of units of the underlying that will be delivered, and where the future delivery will occur. These are important points but relatively minor in this discussion, so they can be left out of the definition to keep it uncluttered.

As noted earlier, a forward contract is a commitment. Each party agrees that it will fulfill its responsibility at the designated future date. Failure to do so constitutes a default and the non-defaulting party can institute legal proceedings to enforce performance. It is important to recognize that although either party could default to the other, only one party at a time can default. The party owing the greater amount could default to the other, but the party owing the lesser amount cannot default because its claim on the other party is greater. The amount owed is always based on the net owed by one party to the other.

To gain a better understanding of forward contracts, it is necessary to examine their payoffs. As noted, forward contracts—and indeed all derivatives—take (derive) their payoffs from the performance of the underlying asset. To illustrate the payoff of a forward contract, start with the assumption that we are at time $t = 0$ and that the forward contract expires at a later date, time $t = T$.[4] The spot price of the underlying asset at time 0 is S_0 and at time T is S_T. Of course, when we initiate the contract at time 0, we do not know what S_T will ultimately be. Remember that the two parties, the buyer and the seller, are going long and short, respectively.

At time $t = 0$, the long and the short agree that the short will deliver the asset to the long at time T for a price of $F_0(T)$. The notation $F_0(T)$ denotes that this value is established at time 0 and applies to a contract expiring at time T. $F_0(T)$ is the forward price.

So, let us assume that the buyer enters into the forward contract with the seller for a price of $F_0(T)$, with delivery of one unit of the underlying asset to occur at time T. Now, let us roll forward to time T, when the price of the underlying is S_T. The long is obligated to pay $F_0(T)$, for which he receives an asset worth S_T. If $S_T > F_0(T)$, it is clear that the transaction has worked out well for the long. He paid $F_0(T)$ and receives something of greater value. Thus, the contract effectively pays off $S_T - F_0(T)$ to the

4 Such notations as $t = 0$ and $t = T$ are commonly used in explaining derivatives. To indicate that $t = 0$ simply means that we initiate a contract at an imaginary time designated like a counter starting at zero. To indicate that the contract expires at $t = T$ simply means that at some future time, designated as T, the contract expires. Time T could be a certain number of days from now or a fraction of a year later or T years later. We will be more specific in later readings that involve calculations. For now, just assume that $t = 0$ and $t = T$ are two dates—the initiation and the expiration—of the contract.

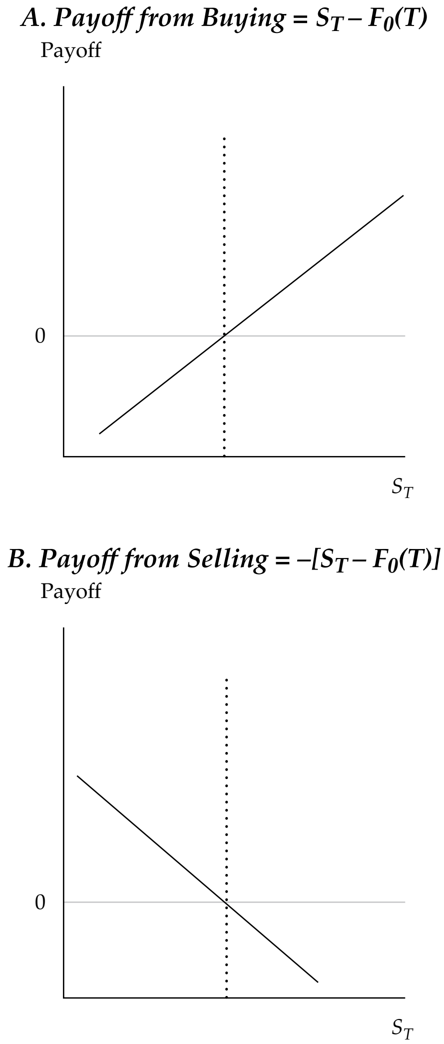
long, which is the value of the contract at expiration. The short has the mirror image of the long. He is required to deliver the asset worth S_T and accept a smaller amount, $F_0(T)$. The contract has a payoff for him of $F_0(T) - S_T$, which is negative. Even if the asset's value, S_T, is less than the forward price, $F_0(T)$, the payoffs are still $S_T - F_0(T)$ for the long and $F_0(T) - S_T$ for the short. We can consolidate these results by writing the short's payoff as the negative of the long's, $-[S_T - F_0(T)]$, which serves as a useful reminder that the long and the short are engaged in a zero-sum game, which is a type of competition in which one participant's gains are the other's losses. Although both lose a modest amount in the sense of both having some costs to engage in the transaction, these costs are relatively small and worth ignoring for our purposes at this time. In addition, it is worthwhile to note how derivatives transform the performance of the underlying. The gain from owning the underlying would be $S_T - S_0$, whereas the gain from owning the forward contract would be $S_T - F_0(T)$. Both figures are driven by S_T, the price of the underlying at expiration, but they are not the same.

For an example, a buyer enters a forward contract to buy gold at a price of $F_0(T)$ = \$1,312.90 per ounce four months from now. The spot price of gold is S_0 = \$1,207.40 per ounce. Four months in the future, the price of the underlying gold is S_T = \$1,275.90 per ounce. The buyer's gain from the forward contract, the payoff from the contract, is the value of gold (at maturity) less the forward price: $S_T - F_0(T)$ = 1,275.90 – 1,312.90 = –\$37.00 per ounce. Because the value of gold when the contract matures is less than the forward price, $S_T < F_0(T)$, the buyer has incurred a loss. Notably, the forward contract seller has a contract payoff, +\$37.00, that is the negative of that of the contract buyer. The gain on owning the underlying, which is $S_T - S_0$ = 1,275.90 – 1,207.40 = \$68.40, differs from the gain (–\$37.00) on the forward contract.

The buyer also enters a forward contract to buy oil at a price of $F_0(T)$ = \$71.86 per barrel four months from now. The spot price of oil is S_0 = \$71.11 per barrel. Four months in the future, the price of the underlying oil is S_T = \$80.96 per barrel. The buyer's gain from the forward contract, the payoff from the contract, is the value of oil less the forward price: $S_T - F_0(T)$ = 80.96 – 71.86 = \$9.10 per barrel. Unlike the forward contract on gold above, because the value of oil when the contract matures is greater than the forward price, $S_T > F_0(T)$, the buyer of the forward contract realizes a gain.

Exhibit 1 illustrates the payoffs from both buying and selling a forward contract.

Exhibit 1 Payoffs from a Forward Contract



The long hopes the price of the underlying will rise above the forward price, $F_0(T)$, whereas the short hopes the price of the underlying will fall below the forward price. Except in the extremely rare event that the underlying price at T equals the forward price, there will ultimately be a winner and a loser.

An important element of forward contracts is that no money changes hands between parties when the contract is initiated. Unlike in the purchase and sale of an asset, there is no value exchanged at the start. The buyer does not pay the seller some money and obtain something. In fact, forward contracts have zero value at the start. They are neither assets nor liabilities. As you will learn in later readings, their values will deviate from zero later as prices move. Forward contracts will almost always have non-zero values at expiration.

As noted previously, the primary purpose of derivatives is for risk management. Although the uses of forward contracts are covered in depth later in the curriculum, there are a few things to note here about the purposes of forward contracts. It should be apparent that locking in the future buying or selling price of an underlying asset can be extremely attractive for some parties. For example, an airline anticipating the purchase of jet fuel at a later date can enter into a forward contract to buy the fuel at a price agreed upon when the contract is initiated. In so doing, the airline has hedged its cost of fuel. Thus, forward contracts can be structured to create a perfect hedge, providing an assurance that the underlying asset can be bought or sold at a price known when the contract is initiated. Likewise, speculators, who ultimately assume the risk laid off by hedgers, can make bets on the direction of the underlying asset without having to invest the money to purchase the asset itself.

Finally, forward contracts need not specifically settle by delivery of the underlying asset. They can settle by an exchange of cash. These contracts—called **non-deliverable forwards** (NDFs), **cash-settled forwards**, or **contracts for differences**—have the same economic effect as do their delivery-based counterparts. For example, for a physical delivery contract, if the long pays $F_0(T)$ and receives an asset worth S_T, the contract is worth $S_T - F_0(T)$ to the long at expiration. A non-deliverable forward contract would have the short simply pay cash to the long in the amount of $S_T - F_0(T)$. The long would not take possession of the underlying asset, but if he wanted the asset, he could purchase it in the market for its current price of S_T. Because he received a cash settlement in the amount of $S_T - F_0(T)$, in buying the asset the long would have to pay out only $S_T - [S_T - F_0(T)]$, which equals $F_0(T)$. Thus, the long could acquire the asset, effectively paying $F_0(T)$, exactly as the contract promised. Transaction costs do make cash settlement different from physical delivery, but this point is relatively minor and can be disregarded for our purposes here.

As previously mentioned, forward contracts are OTC contracts. There is no formal forward contract exchange. Nonetheless, there are exchange-traded variants of forward contracts, which are called futures contracts or just futures.

4 TYPES OF DERIVATIVES: FUTURES

b contrast forward commitments with contingent claims;

c define forward contracts, futures contracts, options (calls and puts), swaps, and credit derivatives and compare their basic characteristics;

Futures contracts are specialized versions of forward contracts that have been standardized and that trade on a futures exchange. By standardizing these contracts and creating an organized market with rules, regulations, and a central clearing facility, the futures markets offer an element of liquidity and protection against loss by default.

Formally, a futures contract is defined as follows:

> *A futures contract is a standardized derivative contract created and traded on a futures exchange in which two parties agree that one party, the buyer, will purchase an underlying asset from the other party, the seller, at a later date and at a price agreed on by the two parties when the contract is initiated and in which there is a daily settling of gains and losses and a credit guarantee by the futures exchange through its clearinghouse.*

First, let us review what standardization means. Recall that in forward contracts, the parties customize the contract by specifying the underlying asset, the time to expiration, the delivery and settlement conditions, and the quantity of the underlying, all according to whatever terms they agree on. These contracts are not traded on an exchange. As noted, the regulation of OTC derivatives markets is increasing, but these contracts are not subject to the traditionally high degree of regulation that applies to securities and futures markets. Futures contracts first require the existence of a futures exchange, a legally recognized entity that provides a market for trading these contracts. Futures exchanges are highly regulated at the national level in all countries. These exchanges specify that only certain contracts are authorized for trading. These contracts have specific underlying assets, times to expiration, delivery and settlement conditions, and quantities. The exchange offers a facility in the form of a physical location and/or an electronic system as well as liquidity provided by authorized market makers.

Probably the most important distinctive characteristic of futures contracts is the daily settlement of gains and losses and the associated credit guarantee provided by the exchange through its clearinghouse. When a party buys a futures contract, it commits to purchase the underlying asset at a later date and at a price agreed upon when the contract is initiated. The counterparty (the seller) makes the opposite commitment, an agreement to sell the underlying asset at a later date and at a price agreed upon when the contract is initiated. The agreed-upon price is called the **futures price**. Identical contracts trade on an ongoing basis at different prices, reflecting the passage of time and the arrival of new information to the market. Thus, as the futures price changes, the parties make and lose money. Rising (falling) prices, of course, benefit (hurt) the long and hurt (benefit) the short. At the end of each day, the clearinghouse engages in a practice called **mark to market**, also known as the **daily settlement**. The clearinghouse determines an average of the final futures trades of the day and designates that price as the **settlement price**. All contracts are then said to be *marked to the settlement price*. For example, if the long purchases the contract during the day at a futures price of £120 and the settlement price at the end of the day is £122, the long's account would be marked for a gain of £2. In other words, the long has made a profit of £2 and that amount is credited to his account, with the money coming from the account of the short, who has lost £2. Naturally, if the futures price decreases, the long loses money and is charged with that loss, and the money is transferred to the account of the short.[5]

The account is specifically referred to as a **margin** account. Of course, in equity markets, margin accounts are commonly used, but there are significant differences between futures margin accounts and equity margin accounts. Equity margin accounts involve the extension of credit. An investor deposits part of the cost of the stock and borrows the remainder at a rate of interest. With futures margin accounts, both parties deposit a required minimum sum of money, but the remainder of the price is not borrowed. This required margin is typically less than 10% of the futures price, which is considerably less than in equity margin trading. In the example above, let us assume that the required margin is £10, which is referred to as the **initial margin**. Both the long and the short put that amount into their respective margin accounts. This money is deposited there to support the trade, not as a form of equity, with the remaining amount borrowed. There is no formal loan created as in equity markets. A futures margin is more of a performance bond or good faith deposit, terms that were previously mentioned. It is simply an amount of money put into an account that covers possible future losses.

Associated with each initial margin is another figure called the **maintenance margin**. The maintenance margin is the amount of money that each participant must maintain in the account after the trade is initiated, and it is always significantly lower than the initial margin. Let us assume that the maintenance margin in this example is £6. If the buyer's account is marked to market with a credit of £2, his margin balance moves to £12, while the seller's account is charged £2 and his balance moves to £8. The clearinghouse then compares each participant's balance with the maintenance margin. At this point, both participants more than meet the maintenance margin.

5 The actual amount of money charged and credited depends on the contract size and the number of contracts. A price of £120 might actually refer to a contract that has a standard size of £100,000. Thus, £120 might actually mean 120% of the standard size, or £120,000. In addition, the parties are likely to hold more than one contract. Hence, the gain of £2 referred to in the text might really mean £2,000 (122% minus 120% times the £100,000 standard size) times the number of contracts held by the party.

Let us say, however, that the price continues to move in the long's favor and, therefore, against the short. A few days later, assume that the short's balance falls to £4, which is below the maintenance margin requirement of £6. The short will then get a **margin call**, which is a request to deposit additional funds. The amount that the short has to deposit, however, is *not* the £2 that would bring his balance up to the maintenance margin. Instead, the short must deposit enough funds to bring the balance up to the initial margin. So, the short must come up with £6. The purpose of this rule is to get the party's position significantly above the minimum level and provide some breathing room. If the balance were brought up only to the maintenance level, there would likely be another margin call soon. A party can choose not to deposit additional funds, in which case the party would be required to close out the contract as soon as possible and would be responsible for any additional losses until the position is closed.

As with forward contracts, neither party pays any money to the other when the contract is initiated. Value accrues as the futures price changes, but at the end of each day, the mark-to-market process settles the gains and losses, effectively resetting the value for each party to zero.

The clearinghouse moves money between the participants, crediting gains to the winners and charging losses to the losers. By doing this on a daily basis, the gains and losses are typically quite small, and the margin balances help ensure that the clearinghouse will collect from the party losing money. As an extra precaution, in fast-moving markets, the clearinghouse can make margin calls during the day, not just at the end of the day. Yet there still remains the possibility that a party could default. A large loss could occur quickly and consume the entire margin balance, with additional money owed.[6] If the losing party cannot pay, the clearinghouse provides a guarantee that it will make up the loss, which it does by maintaining an insurance fund. If that fund were depleted, the clearinghouse could levy a tax on the other market participants, though that has never happened.

Some futures contracts contain a provision limiting price changes. These rules, called **price limits**, establish a band relative to the previous day's settlement price, within which all trades must occur. If market participants wish to trade at a price above the upper band, trading stops, which is called **limit up**, until two parties agree on a trade at a price lower than the upper limit. Likewise, if market participants wish to trade at a price below the lower band, which is called **limit down**, no trade can take place until two parties agree to trade at a price above the lower limit. When the market hits these limits and trading stops, it is called **locked limit**. Typically, the exchange rules provide for an expansion of the limits the next day. These price limits, which may be somewhat objectionable to proponents of free markets, are important in helping the clearinghouse manage its credit exposure. Just because two parties wish to trade a futures contract at a price beyond the limits does not mean they should be allowed to do so. The clearinghouse is a third participant in the contract, guaranteeing to each party that it ensures against the other party defaulting. Therefore, the clearinghouse has a vested interest in the price and considerable exposure. Sharply moving prices make it more difficult for the clearinghouse to collect from the parties losing money.

Most participants in futures markets buy and sell contracts, collecting their profits and incurring their losses, with no ultimate intent to make or take delivery of the underlying asset. For example, the long may ultimately sell her position before expiration. When a party re-enters the market at a later date but before expiration and

---

6 For example, let us go back to when the short had a balance of £4, which is £2 below the maintenance margin and £6 below the initial margin. The short will get a margin call, but suppose he elects not to deposit additional funds and requests that his position be terminated. In a fast-moving market, the price might increase more than £4 before his broker can close his position. The remaining balance of £4 would then be depleted, and the short would be responsible for any additional losses.

engages in the opposite transaction—a long selling her previously opened contract or a short buying her previously opened contract—the transaction is referred to as an offset. The clearinghouse marks the contract to the current price relative to the previous settlement price and closes out the participant's position.

At any given time, the number of outstanding contracts is called the **open interest**. Each contract counted in the open interest has a long and a corresponding short. The open interest figure changes daily as some parties open up new positions, while other parties offset their old positions. It is theoretically possible that all longs and shorts offset their positions before expiration, leaving no open interest when the contract expires, but in practice there is nearly always some open interest at expiration, at which time there is a final delivery or settlement.

When discussing forward contracts, we noted that a contract could be written such that the parties engage in physical delivery or cash settlement at expiration. In the futures markets, the exchange specifies whether physical delivery or cash settlement applies. In physical delivery contracts, the short is required to deliver the underlying asset at a designated location and the long is required to pay for it. Delivery replaces the mark-to-market process on the final day. It also ensures an important principle that you will use later: *The futures price converges to the spot price at expiration.* Because the short delivers the actual asset and the long pays the current spot price for it, the futures price at expiration has to be the spot price at that time. Alternatively, a futures contract initiated right at the instant of expiration is effectively a spot transaction and, therefore, the futures price at expiration must equal the spot price. Following this logic, in cash settlement contracts, there is a final mark to market, with the futures price formally set to the spot price, thereby ensuring automatic convergence.

In discussing forward contracts, we described the process by which they pay off as the spot price at expiration minus the forward price, $S_T - F_0(T)$, the former determined at expiration and the latter agreed upon when the contract is initiated. Futures contracts basically pay off the same way, but there is a slight difference. Let us say the contract is initiated on Day 0 and expires on Day T. The intervening days are designated Days 1, 2, …, T. The initial futures price is designated $f_0(T)$ and the daily settlement prices on Days 1, 2, …, T are designated $f_1(T), f_2(T), …, f_T(T)$. There are, of course, futures prices within each trading day, but let us focus only on the settlement prices for now. For simplicity, let us assume that the long buys at the settlement price on Day 0 and holds the position all the way to expiration. Through the mark-to-market process, the cash flows to the account of the long will be

$f_1(T) - f_0(T)$ on Day 1

$f_2(T) - f_1(T)$ on Day 2

$f_3(T) - f_2(T)$ on Day 3

…

$f_T(T) - f_{T-1}(T)$ on Day T

These add up to

$f_T(T) - f_0(T)$ on Day T.

And because of the convergence of the final futures price to the spot price,

$f_T(T) - f_0(T) = S_T - f_0(T)$,

which is the same as with forward contracts.[7] Note, however, that the timing of these profits is different from that of forwards. Forward contracts realize the full amount, $S_T - f_0(T)$, at expiration, whereas futures contracts realize this amount in parts on a

7 Because of this equivalence, we will not specifically illustrate the profit graphs of futures contracts. You can generally treat them the same as those of forwards, which were shown in Exhibit 1.

day-to-day basis. Naturally, the time value of money principle says that these are not equivalent amounts of money. But the differences tend to be small, particularly in low-interest-rate environments, some of these amounts are gains and some are losses, and most futures contracts have maturities of less than a year.

But the near equivalence of the profits from a futures and a forward contract disguises an important distinction between these types of contracts. In a forward contact, with the entire payoff made at expiration, a loss by one party can be large enough to trigger a default. Hence, forward contracts are subject to default and require careful consideration of the credit quality of the counterparties. Because futures contracts settle gains and collect losses daily, the amounts that could be lost upon default are much smaller and naturally give the clearinghouse much greater flexibility to manage the credit risk it assumes.

Unlike forward markets, futures markets are highly regulated at the national level. National regulators are required to approve new futures exchanges and even new contracts proposed by existing exchanges as well as changes in margin requirements, price limits, and any significant changes in trading procedures. Violations of futures regulations can be subject to governmental prosecution. In addition, futures markets are far more transparent than forward markets. Futures prices, volume, and open interest are widely reported and easily obtained. Futures prices of nearby expiring contracts are often used as proxies for spot prices, particularly in decentralized spot markets, such as gold, which trades in spot markets all over the world.

In spite of the advantages of futures markets over forward markets, forward markets also have advantages over futures markets. Transparency is not always a good thing. Forward markets offer more privacy and fewer regulatory encumbrances. In addition, forward markets offer more flexibility. With the ability to tailor contracts to the specific needs of participants, forward contracts can be written exactly the way the parties want. In contrast, the standardization of futures contracts makes it more difficult for participants to get exactly what they want, even though they may get close substitutes. Yet, futures markets offer a valuable credit guarantee.

Like forward markets, futures markets can be used for hedging or speculation. For example, a jewelry manufacturer can buy gold futures, thereby hedging the price it will have to pay for one of its key inputs. Although it is more difficult to construct a futures strategy that hedges perfectly than to construct a forward strategy that does so, futures offer the benefit of the credit guarantee. It is not possible to argue that futures are better than forwards or vice versa. Market participants always trade off advantages against disadvantages. Some participants prefer futures, and some prefer forwards. Some prefer one over the other for certain risks and the other for other risks. Some might use one for a particular risk at a point in time and a different instrument for the same risk at another point in time. The choice is a matter of taste and constraints.

The third and final type of forward commitment we will cover is swaps. They go a step further in committing the parties to buy and sell something at a later date: They obligate the parties to a sequence of multiple purchases and sales.

5 TYPES OF DERIVATIVES: SWAPS

b contrast forward commitments with contingent claims;

c define forward contracts, futures contracts, options (calls and puts), swaps, and credit derivatives and compare their basic characteristics;

The concept of a swap is that two parties exchange (swap) a series of cash flows. One set of cash flows is variable or floating and will be determined by the movement of an underlying asset or rate. The other set of cash flows can be variable and determined by a different underlying asset or rate, or it can be fixed. Formally, a swap is defined as follows:

> *A swap is an over-the-counter derivative contract in which two parties agree to exchange a series of cash flows whereby one party pays a variable series that will be determined by an underlying asset or rate and the other party pays either (1) a variable series determined by a different underlying asset or rate or (2) a fixed series.*

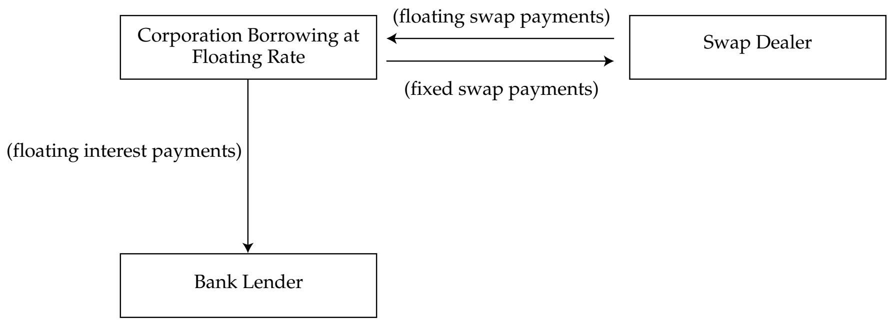
As with forward contracts, swap contracts also contain other terms—such as the identity of the underlying, the relevant payment dates, and the payment procedure—that are negotiated between the parties and written into the contract. A swap is a bit more like a forward contract than a futures contract in that it is an OTC contract, so it is privately negotiated and subject to default. Nonetheless, the similarities between futures and forwards apply to futures and swaps and, indeed, combinations of futures contracts expiring at different dates are often compared to swaps.

As with forward contracts, either party can default but only one party can default at a particular time. The money owed is always based on the net owed by one party to the other. Hence, the party owing the lesser amount cannot default to the party owing the greater amount. Only the latter can default, and the amount it owes is the net of what it owes and what is owed to it, which is also true with forwards.

Swaps are relatively young financial instruments, having been created only in the early 1980s. Thus, it may be somewhat surprising to learn that the swap is the most widely used derivative, a likely result of its simplicity and embracement by the corporate world. The most common swap is the **fixed-for-floating interest rate swap**. In fact, this type of swap is so common that it is often called a "plain vanilla swap" or just a "vanilla swap," owing to the notion that vanilla ice cream is considered plain (albeit tasty).

Let us examine a scenario in which the vanilla interest rate swap is frequently used. Suppose a corporation borrows from a bank at a floating rate. It would prefer a fixed rate, which would enable it to better anticipate its cash flow needs in making its interest payments.[8] The corporation can effectively convert its floating-rate loan to a fixed-rate loan by adding a swap, as shown in Exhibit 2.

Exhibit 2 Using an Interest Rate Swap to Convert a Floating-Rate Loan to a Fixed-Rate Loan



8 Banks prefer to make floating-rate loans because their own funding is typically short term and at floating rates. Thus, their borrowing rates reset frequently, giving them a strong incentive to pass that risk on to their customers through floating-rate loans.

The interest payments on the loan are tied to a specific floating rate. For a dollar-based loan, that rate has historically been US dollar Libor.[9] The payments would be based on the rate from the Libor market on a specified reset date times the loan balance times a factor reflecting the number of days in the current interest calculation period. The actual payment is made at a later date. Thus, for a loan balance of, say, \$10 million with monthly payments, the rate might be based on Libor on the first business day of the month, with interest payable on the first business day of the next month, which is the next reset date, and calculated as \$10 million times the rate times 30/360. The 30/360 convention, an implicit assumption of 30 days in a month, is common but only one of many interest calculation conventions used in the financial world. Often, "30" is replaced by the exact number of days since the last interest payment. The use of a 360-day year is a common assumption in the financial world, which originated in the pre-calculator days when an interest rate could be multiplied by a number like 30/360, 60/360, 90/360, etc., more easily than if 365 were used.

Whatever the terms of the loan are, the terms of the swap are typically set to match those of the loan. Thus, a Libor-based loan with monthly payments based on the 30/360 convention would be matched with a swap with monthly payments based on Libor and the 30/360 convention and the same reset and payment dates. Although the loan has an actual balance (the amount owed by borrower to creditor), the swap does not have such a balance owed by one party to the other. Thus, it has no principal, but it does have a balance of sorts, called the **notional principal**, which ordinarily matches the loan balance. A loan with only one principal payment, the final one, will be matched with a swap with a fixed notional principal. An amortizing loan, which has a declining principal balance, will be matched with a swap with a pre-specified declining notional principal that matches the loan balance.

As with futures and forwards, no money changes hands at the start; thus, the value of a swap when initiated must be zero. The fixed rate on the swap is determined by a process that forces the value to zero, a procedure that will be covered later in the curriculum. As market conditions change, the value of a swap will deviate from zero, being positive to one party and negative to the other.

As with forward contracts, swaps are subject to default, but because the notional amount of a swap is not typically exchanged, the credit risk of a swap is much less than that of a loan.[10] The only money passing from one party to the other is the net difference between the fixed and floating interest payments. In fact, the parties do not even pay each other. Only one party pays the other, as determined by the net of the greater amount owed minus the lesser amount. This does not mean that swaps are not subject to a potentially large amount of credit risk. At a given point in time, one party could default, effectively owing the value of all remaining payments, which could substantially exceed the value that the non-defaulting party owes to the defaulting party. Thus, there is indeed credit risk in a swap. This risk must be managed by careful analysis before the transaction and by the potential use of such risk-mitigating measures as collateral.

There are also interest rate swaps in which one party pays on the basis of one interest rate and the other party pays on the basis of a different interest rate. For example, one party might make payments at Libor, whereas the other might make payments on the

---

9 Libor is being phased out, as the panel of banks will no longer be required to submit quotations after 2021. In anticipation of this, market participants and regulators have been working to develop alternative reference rates.

10 It is possible that the notional principal will be exchanged in a currency swap, whereby each party makes a series of payments to the other in different currencies. Whether the notional principal is exchanged depends on the purpose of the swap. This point will be covered later in the curriculum. At this time, you should see that it would be fruitless to exchange notional principals in an interest rate swap because that would mean each party would give the other the same amount of money when the transaction is initiated and re-exchange the same amount of money when the contract terminates.

basis of the U. S. Treasury bill rate. The difference between Libor and the T-bill rate, often called the TED spread (T-bills versus Eurodollar), is a measure of the credit risk premium of London banks, which have historically borrowed short term at Libor, versus that of the U. S. government, which borrows short term at the T-bill rate. This transaction is called a basis swap. There are also swaps in which the floating rate is set as an average rate over the period, in accordance with the convention for many loans. Some swaps, called overnight indexed swaps, are tied to a Fed funds–type rate, reflecting the rate at which banks borrow overnight. As we will cover later, there are many other different types of swaps that are used for a variety of purposes. The plain vanilla swap is merely the simplest and most widely used.

Because swaps, forwards, and futures are forward commitments, they can all accomplish the same thing. One could create a series of forwards or futures expiring at a set of dates that would serve the same purpose as a swap. Although swaps are better suited for risks that involve multiple payments, at its most fundamental level, a swap is more or less just a series of forwards and, acknowledging the slight differences discussed above, more or less just a series of futures.

EXAMPLE 3

Forward Contracts, Futures Contracts, and Swaps

1 Which of the following characterizes forward contracts and swaps but **not** futures?
   - A They are customized.
   - B They are subject to daily price limits.
   - C Their payoffs are received on a daily basis.

2 Which of the following distinguishes forwards from swaps?
   - A Forwards are OTC instruments, whereas swaps are exchange traded.
   - B Forwards are regulated as futures, whereas swaps are regulated as securities.
   - C Swaps have multiple payments, whereas forwards have only a single payment.

3 Which of the following occurs in the daily settlement of futures contracts?
   - A Initial margin deposits are refunded to the two parties.
   - B Gains and losses are reported to other market participants.
   - C Losses are charged to one party and gains credited to the other.

Solution to 1:

A is correct. Forwards and swaps are OTC contracts and, therefore, are customized. Futures are exchange traded and, therefore, are standardized. Some futures contracts are subject to daily price limits and their payoffs are received daily, but these characteristics are not true for forwards and swaps.

Solution to 2:

C is correct. Forwards and swaps are OTC instruments and both are regulated as such. Neither is regulated as a futures contract or a security. A swap is a series of multiple payments at scheduled dates, whereas a forward has only one payment, made at its expiration date.

Solution to 3:

C is correct. Losses and gains are collected and distributed to the respective parties. There is no specific reporting of these gains and losses to anyone else. Initial margin deposits are not refunded and, in fact, additional deposits may be required.

This material completes our introduction to forward commitments. All forward commitments are firm contracts. The parties are required to fulfill the obligations they agreed to. The benefit of this rigidity is that neither party pays anything to the other when the contract is initiated. If one party needs some flexibility, however, it can get it by agreeing to pay the other party some money when the contract is initiated. When the contract expires, the party who paid at the start has some flexibility in deciding whether to buy the underlying asset at the fixed price. Thus, that party did not actually agree to do anything. It had a choice. This is the nature of contingent claims.

6 CONTINGENT CLAIMS: OPTIONS

- **b** contrast forward commitments with contingent claims;
- **c** define forward contracts, futures contracts, options (calls and puts), swaps, and credit derivatives and compare their basic characteristics;
- **d** determine the value at expiration and profit from a long or a short position in a call or put option;

A **contingent claim** is a derivative in which the outcome or payoff is dependent on the outcome or payoff of an underlying asset. Although this characteristic is also associated with forward commitments, a contingent claim has come to be associated with a *right*, but not an *obligation*, to make a final payment contingent on the performance of the underlying. Given that the holder of the contingent claim has a choice, the term *contingent claim* has become synonymous with the term *option*. The holder has a choice of whether or not to exercise the option. This choice creates a payoff that transforms the underlying payoff in a more pronounced manner than does a forward, futures, or swap. Those instruments provide linear payoffs: As the underlying goes up (down), the derivative gains (loses). The further up (down) the underlying goes, the more the derivative gains (loses). Options are different in that they limit losses in one direction. In addition, options can pay off as the underlying goes down. Hence, they transform the payoffs of the underlying into something quite different.

6.1 Options

We might say that an option, as a contingent claim, grants the right but not the obligation to buy an asset at a later date and at a price agreed on when the option is initiated. But there are so many variations of options that we cannot settle on this statement as a good formal definition. For one thing, options can also grant the right to sell instead of the right to buy. Moreover, they can grant the right to buy or sell earlier than at expiration. So, let us see whether we can combine these points into an all-encompassing definition of an option.

> *An option is a derivative contract in which one party, the buyer, pays a sum of money to the other party, the seller or writer, and receives the right to either buy or sell an underlying asset at a fixed price either on a specific expiration date or at any time prior to the expiration date.*

Unfortunately, even that definition does not cover every unique aspect of options. For example, options can be created in the OTC market and customized to the terms of each party, or they can be created and traded on options exchanges and standardized. As with forward contracts and swaps, customized options are subject to default, are less regulated, and are less transparent than exchange-traded derivatives. Exchange-traded options are protected against default by the clearinghouse of the options exchange and are relatively transparent and regulated at the national level. As noted in the definition above, options can be terminated early or at their expirations. When an option is terminated, either early or at expiration, the holder of the option chooses whether to exercise it. If he exercises it, he either buys or sells the underlying asset, but he does not have both rights. The right to buy is one type of option, referred to as a **call** or **call option**, whereas the right to sell is another type of option, referred to as a **put** or **put option**. With one very unusual and advanced exception that we do not cover, an option is either a call or a put, and that point is made clear in the contract.

An option is also designated as exercisable early (before expiration) or only at expiration. Options that can be exercised early are referred to as **American-style**. Options that can be exercised only at expiration are referred to as **European-style**. *It is extremely important that you do not associate these terms with where these options are traded*. Both types of options trade on all continents.

As with forwards and futures, an option can be exercised by physical delivery or cash settlement, as written in the contract. For a call option with physical delivery, upon exercise the underlying asset is delivered to the call buyer, who pays the call seller the exercise price. For a put option with physical delivery, upon exercise the put buyer delivers the underlying asset to the put seller and receives the strike price. For a cash settlement option, exercise results in the seller paying the buyer the cash equivalent value as if the asset were delivered and paid for.

The fixed price at which the underlying asset can be purchased is called the **exercise price** (also called the "strike price," the "strike," or the "striking price"). This price is somewhat analogous to the forward price because it represents the price at which the underlying will be purchased or sold if the option is exercised. The forward price, however, is set in the pricing of the contract such that the contract value at the start is zero. The strike price of the option is chosen by the participants. The actual price or value of the option is an altogether different concept.

As noted, the buyer pays the writer a sum of money called the **option premium**, or just the "premium." It represents a fair price of the option, and in a well-functioning market, it would be the value of the option. Consistent with everything we know about finance, it is the present value of the cash flows that are expected to be received by the holder of the option during the life of the option. At this point, we will not get into how this price is determined, but you will learn that later. For now, there are some fundamental concepts you need to understand, which form a basis for understanding how options are priced and why anyone would use an option.

Because the option buyer (the long) does not have to exercise the option, beyond the initial payment of the premium, there is no obligation of the long to the short. Thus, only the short can default, which would occur if the long exercises the option and the short fails to do what it is supposed to do. Thus, in contrast to forwards and swaps, in which either party could default to the other, default in options is possible only from the short to the long.

Ruling out the possibility of default for now, let us examine what happens when an option expires. Using the same notation used previously, let S_T be the price of the underlying at the expiration date, T, and X be the exercise price of the option. Remember that a call option allows the holder, or long, to pay X and receive the underlying. It should be obvious that the long would exercise the option at expiration if S_T is greater than X, meaning that the underlying value is greater than what he would pay to obtain the underlying. Otherwise, he would simply let the option expire. Thus, on the expiration date, the option is described as having a payoff of $Max(0,S_T - X)$.

Because the holder of the option would be entitled to exercise it and claim this amount, it also represents the value of the option at expiration. Let us denote that value as c_T. Thus,

$$c_T = Max(0,S_T - X) \qquad \text{(payoff to the call buyer),}$$

which is read as "take the maximum of either zero or $S_T - X$." For example, suppose you buy a call option with an exercise price of 50 and an expiration of three months for a premium of 1.50 when the stock is trading at 45. At expiration, consider the outcomes when the stock's price is 45, 50, or 55. The buyer's payoffs would be:

For $S_T = 45$, payoff $= c_T = Max(0,S_T - X) = Max(0,45 - 50) = \text{Max}(0, -5) = 0$.

For $S_T = 50$, payoff $= c_T = Max(0,S_T - X) = Max(0,50 - 50) = \text{Max}(0, 0) = 0$.

For $S_T = 55$, payoff $= c_T = Max(0,S_T - X) = Max(0,55 - 50) = \text{Max}(0, 5) = 5$.

Thus, if the underlying value exceeds the exercise price ($S_T > X$), then the option value is positive and equal to $S_T - X$. The call option is then said to be **in the money**. If the underlying value is less than the exercise price ($S_T < X$), then $S_T - X$ is negative; zero is greater than a negative number, so the option value would be zero. When the underlying value is less than the exercise price, the call option is said to be **out of the money**. When $S_T = X$, the call option is said to be **at the money**, although at the money is, for all practical purposes, out of the money because the value is still zero.

This payoff amount is also the value of the option at expiration. It represents value because it is what the option is worth at that point. If the holder of the option sells it to someone else an instant before expiration, it should sell for that amount because the new owner would exercise it and capture that amount. To the seller, the value of the option at that point is $-Max(0,S_T - X)$, which is negative to the seller if the option is in the money and zero otherwise.

Using the payoff value and the price paid for the option, we can determine the profit from the strategy, which is denoted with the Greek symbol Π. Let us say the buyer paid c_0 for the option at time 0. Then the profit is

$$\Pi = Max(0,S_T - X) - c_0 \qquad \text{(profit to the call buyer),}$$

Continuing with the example with underlying prices at expiration of 45, 50, or 55, the call buyer's profit would be:

For $S_T = 45$, profit $= Max(0,S_T - X) - c_0 = Max(0,45 - 50) - 1.50 = -1.50$.

For $S_T = 50$, profit $= Max(0,S_T - X) - c_0 = Max(0,50 - 50) - 1.50 = -1.50$.

For $S_T = 55$, profit $= Max(0,S_T - X) - c_0 = Max(0,55 - 50) - 1.50 = 3.50$.

To the seller, who received the premium at the start, the payoff is

$$-c_T = -Max(0,S_T - X) \qquad \text{(payoff to the call seller),}$$

At expiration, the call seller's payoffs are:

For $S_T = 45$, payoff $= -c_T = -Max(0,S_T - X) = -Max(0,45 - 50) = 0$.

For $S_T = 50$, payoff $= -c_T = -Max(0,S_T - X) = -Max(0,50 - 50) = 0$.

For $S_T = 55$, payoff $= -c_T = -Max(0,S_T - X) = -Max(0,55 - 50) = -5$.

The call seller's profit is

$\Pi = -Max(0,S_T - X) + c_0$ (profit to the call seller),

Finally, at expiration, the call seller's profit for each underlying price at expiration are:

For $S_T = 45$, profit $= -Max(0,S_T - X) + c_0 = -Max(0,45 - 50) + 1.50 = 1.50$.

For $S_T = 50$, profit $= -Max(0,S_T - X) + c_0 = -Max(0,50 - 50) + 1.50 = 1.50$.

For $S_T = 55$, profit $= -Max(0,S_T - X) + c_0 = -Max(0,55 - 50) + 1.50 = -3.50$.

For any given price at expiration, the call seller's payoff or profit is equal to the negative of the call buyer's payoff or profit.

EXAMPLE 4

Call Option Payoffs and Profit at Expiration

Consider a call option selling for \$7 in which the exercise price is \$100 and the price of the underlying is \$98.

1 Determine the value at expiration and the profit for a call buyer under the following outcomes:

A The price of the underlying at expiration is \$102.

B The price of the underlying at expiration is \$94.

2 Determine the value at expiration and the profit for a call seller under the following outcomes:

A The price of the underlying at expiration is \$91.

B The price of the underlying at expiration is \$101.

Solution to 1:

A If the price of the underlying at expiration is \$102,

$$\begin{aligned}\text{The call buyer's value at expiration} &= c_T = Max(0,S_T - X) \\ &= Max(0,102 - 100) = \$2. \\ \text{The call buyer's profit} &= \Pi = c_T - c_0 = 2 - 7 = -\$5.\end{aligned}$$

B If the price of the underlying at expiration is \$94,

$$\begin{aligned}\text{The call buyer's value at expiration} &= c_T = Max(0,S_T - X) \\ &= Max(0,94 - 100) = \$0. \\ \text{The call buyer's profit} &= \Pi = c_T - c_0 = 0 - 7 = -\$7.\end{aligned}$$

Solution to 2:

A If the price of the underlying at expiration is \$91,

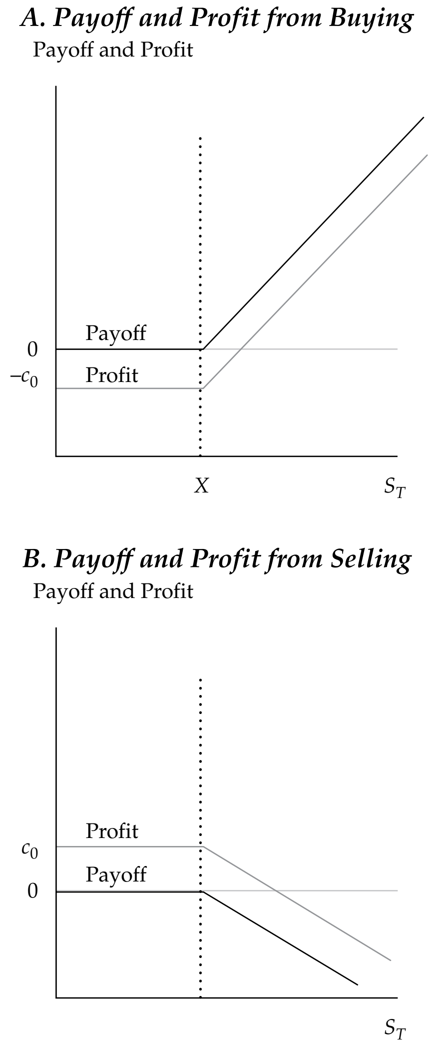
$$\text{The call seller's value at expiration} = -c_T = -Max(0,S_T - X)$$
$$= -Max(0,91 - 100) = \$0.$$
$$\text{The call seller's profit} = \Pi = -c_T + c_0 = 0 + 7 = \$7$$

B If the price of the underlying at expiration is \$101,

$$\text{The call seller's value at expiration} = -c_T = -Max(0,S_T - X)$$
$$= -Max(0,101 - 100) = -\$1.$$
$$\text{The call seller's profit} = \Pi = -c_T + c_0 = -1 + 7 = \$6.$$

Exhibit 3 illustrates the payoffs and profits to the call buyer and seller as graphical representations of these equations, with the payoff or value at expiration indicated by the dark line and the profit indicated by the light line. Note in Panel A that the buyer has no upper limit on the profit and has a fixed downside loss limit equal to the premium paid for the option. Such a condition, with limited loss and unlimited gain, is a temptation to many unsuspecting investors, but keep in mind that the graph does not indicate the frequency with which gains and losses will occur. Panel B is the mirror image of Panel A and shows that the seller has unlimited losses and limited gains. One might suspect that selling a call is, therefore, the worst investment strategy possible. Indeed, it is a risky strategy, but at this point these are only simple strategies. Other strategies can be added to mitigate the seller's risk to a substantial degree.

Exhibit 3 Payoff and Profit from a Call Option



Now let us consider put options. Recall that a put option allows its holder to sell the underlying asset at the exercise price. Thus, the holder should exercise the put at expiration if the underlying asset is worth less than the exercise price ($S_T < X$). In that case, the put is said to be in the money. If the underlying asset is worth the same as the exercise price ($S_T = X$), meaning the put is at the money, or more than the exercise price ($S_T > X$), meaning the put is out of the money, the option holder would not exercise it and it would expire with zero value. Thus, the payoff to the put holder is

$$p_T = Max(0,X - S_T) \quad \text{(payoff to the put buyer),}$$

If the put buyer paid p_0 for the put at time 0, the profit is

$$\Pi = Max(0,X - S_T) - p_0 \quad \text{(profit to the put buyer),}$$

And for the seller, the payoff is

$$-p_T = -Max(0,X - S_T) \quad \text{(payoff to the put seller),}$$

And the profit is

$$\Pi = -Max(0,X - S_T) + p_0 \quad \text{(profit to the put seller),}$$

To illustrate the payoffs and profit to a put buyer and put seller, assume the put price (p_0) is 1.50, the exercise price (X) is 20.00, and the stock price at expiration (S_T) is either 18.00 or 22.00. The put buyer's payoff is:

For S_T = 18, payoff = $p_T = Max(0,X - S_T) = Max(0,20 - 18) = 2$.

For S_T = 22, payoff = $p_T = Max(0,X - S_T) = Max(0,20 - 22) = 0$.

The put buyer's profit is:

For S_T = 18, profit = $Max(0,X - S_T) - p_0 = Max(0,20 - 18) - 1.50 = 0.50$.

For S_T = 22, profit = $Max(0,X - S_T) - p_0 = Max(0,20 - 22) - 1.50 = -1.50$.

The put seller's payoff is:

For S_T = 18, payoff = $-p_T = -Max(0,X - S_T) = -Max(0,20 - 18) = -2$.

For S_T = 22, payoff = $-p_T = -Max(0,X - S_T) = -Max(0,20 - 22) = 0$.

Finally, the put seller's profit is:

For S_T = 18, profit = $-Max(0,X - S_T) + p_0 = -Max(0,20 - 18) + 1.50 = -0.50$.

For S_T = 22, profit = $-Max(0,X - S_T) + p_0 = -Max(0,20 - 22) + 1.50 = 1.50$.

For a given stock price at expiration, the put seller's payoff or profit are the negative of the put buyer's payoff or profit.

EXAMPLE 5

Put Option Payoffs and Profit at Expiration

Consider a put option selling for \$4 in which the exercise price is \$60 and the price of the underlying is \$62.

1 Determine the value at expiration and the profit for a put buyer under the following outcomes:
  - A The price of the underlying at expiration is \$62.
  - B The price of the underlying at expiration is \$55.

2 Determine the value at expiration and the profit for a put seller under the following outcomes:
  - A The price of the underlying at expiration is \$51.
  - B The price of the underlying at expiration is \$68.

Solution to 1:

A If the price of the underlying at expiration is \$62,

The put buyer's value at expiration = $p_T = Max(0,X - S_T)$
$= Max(0,60 - 62) = \$0$.
The put buyer's profit = $\Pi = p_T - p_0 = 0 - 4 = -\$4$.

B If the price of the underlying at expiration is \$55,

The put buyer's value at expiration = $p_T = Max(0,X - S_T)$
$= Max(0,60 - 55) = \$5$.
The put buyer's profit = $\Pi = p_T - p_0 = 5 - 4 = \$1$.

Solution to 2:

A If the price of the underlying at expiration is \$51,

$$\text{The put seller's value at expiration} = -p_T = -Max(0,X - S_T)$$
$$= -Max(0,60 - 51) = -\$9.$$
$$\text{The put seller's profit} = \Pi = -p_T + p_0 = -9 + 4 = -\$5.$$

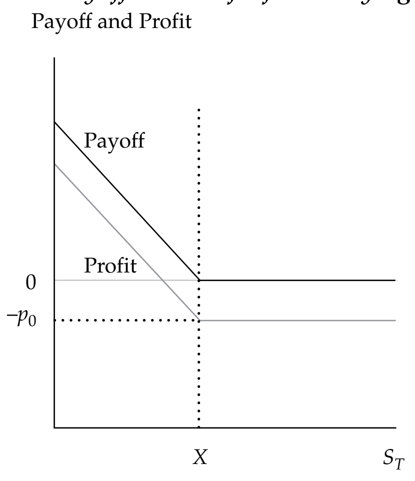
B If the price of the underlying at expiration is \$68,

$$\text{The put seller's value at expiration} = -p_T = -Max(0,X - S_T)$$
$$= -Max(0,60 - 68) = \$0.$$
$$\text{The put seller's profit} = \Pi = -p_T + p_0 = 0 + 4 = \$4.$$

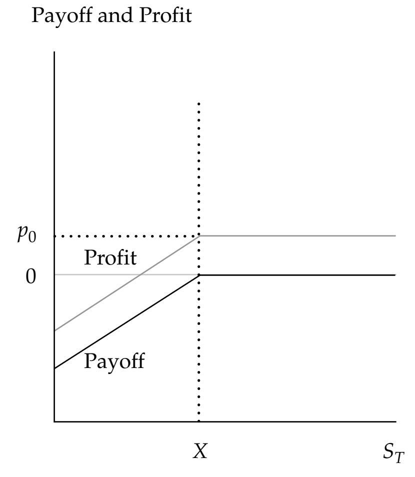
Exhibit 4 illustrates the payoffs and profits to the buyer and seller of a put.

Exhibit 4 Payoff and Profit from a Put Option

A. Payoff and Profit from Buying



B. Payoff and Profit from Selling



The put buyer has a limited loss, and although the gain is limited by the fact that the underlying value cannot go below zero, the put buyer does gain more the lower the value of the underlying. In this manner, we see how a put option is like insurance. Bad outcomes for the underlying trigger a payoff for both the insurance policy and the put, whereas good outcomes result only in loss of the premium. The put seller, like the insurer, has a limited gain and a loss that is larger the lower the value of the underlying. As with call options, these graphs must be considered carefully because they do not indicate the frequency with which gains and losses will occur. At this point, it should be apparent that buying a call option is consistent with a bullish point of view and buying a put option is consistent with a bearish point of view. Moreover, in contrast to forward commitments, which have payoffs that are linearly related to the payoffs of the underlying (note the straight lines in Exhibit 1), contingent claims have payoffs that are non-linear in relation to the underlying. There is linearity over a range—say, from 0 to X or from X upward or downward—but over the entire range of values for the underlying, the payoffs of contingent claims cannot be depicted with a single straight line.

We have seen only a snapshot of the payoff and profit graphs that can be created with options. Calls can be combined with puts, the underlying asset, and other calls or puts with different expirations and exercise prices to create a diverse set of payoff and profit graphs, some of which are covered later in the curriculum.

Before leaving options, let us again contrast the differences between options and forward commitments. With forward commitments, the parties agree to trade an underlying asset at a later date and at a price agreed upon when the contract is initiated. Neither party pays any cash to the other at the start. With options, the buyer pays cash to the seller at the start and receives the right, but not the obligation, to buy (if a call) or sell (if a put) the underlying asset at expiration at a price agreed upon (the exercise price) when the contract is initiated. In contrast to forwards, futures, and swaps, options do have value at the start: the premium paid by buyer to seller. That premium pays for the *right*, eliminating the *obligation*, to trade the underlying at a later date, as would be the case with a forward commitment.

Although there are numerous variations of options, most have the same essential features described here. There is, however, a distinctive family of contingent claims that emerged in the early 1990s and became widely used and, in some cases, heavily criticized. These instruments are known as credit derivatives.

7 CONTINGENT CLAIMS: CREDIT DERIVATIVES

b contrast forward commitments with contingent claims;

c define forward contracts, futures contracts, options (calls and puts), swaps, and credit derivatives and compare their basic characteristics;

Credit risk is surely one of the oldest risks known to mankind. Human beings have been lending things to each other for thousands of years, and even the most primitive human beings must have recognized the risk of lending some of their possessions to their comrades. Until the last 20 years or so, however, the management of credit risk was restricted to simply doing the best analysis possible before making a loan, monitoring the financial condition of the borrower during the loan, limiting the exposure to a given party, and requiring collateral. Some modest forms of insurance against credit risk have existed for a number of years, but insurance can be a slow and cumbersome way of protecting against credit loss. Insurance is typically highly regulated, and insurance laws are usually very consumer oriented. Thus, credit insurance as a financial product has met with only modest success.

In the early 1990s, however, the development of the swaps market led to the creation of derivatives that would hedge credit risk. These instruments came to be known as **credit derivatives**, and they avoided many of the regulatory constraints of the traditional insurance industry. Here is a formal definition:

> *A credit derivative is a class of derivative contracts between two parties, a credit protection buyer and a credit protection seller, in which the latter provides protection to the former against a specific credit loss.*

One of the first credit derivatives was a **total return swap**, in which the underlying is typically a bond or loan, in contrast to, say, a stock or stock index. The credit protection buyer offers to pay the credit protection seller the total return on the underlying bond. This total return consists of all interest and principal paid by the borrower plus any changes in the bond's market value. In return, the credit protection seller typically pays the credit protection buyer either a fixed or a floating rate of interest. Thus, if the bond defaults, the credit protection seller must continue to make its promised payments, while receiving a very small return or virtually no return from the credit protection buyer. If the bond incurs a loss, as it surely will if it defaults, the credit protection seller effectively pays the credit protection buyer.

Another type of credit derivative is the **credit spread option**, in which the underlying is the credit (yield) spread on a bond, which is the difference between the bond's yield and the yield on a benchmark default-free bond. As you will learn in the fixed-income material, the credit spread is a reflection of investors' perception of credit risk. Because a credit spread option requires a credit spread as the underlying, this type of derivative works only with a traded bond that has a quoted price. The credit protection buyer selects the strike spread it desires and pays the option premium to the credit protection seller. At expiration, the parties determine whether the option is in the money by comparing the bond's yield spread with the strike chosen, and if it is, the credit protection seller pays the credit protection buyer the established payoff. Thus, this instrument is essentially a call option in which the underlying is the credit spread.

A third type of credit derivative is the **credit-linked note (CLN)**. With this derivative, the credit protection buyer holds a bond or loan that is subject to default risk (the underlying reference security) and issues its own security (the credit-linked note) with the condition that if the bond or loan it holds defaults, the principal payoff on the credit-linked note is reduced accordingly. Thus, the buyer of the credit-linked note effectively insures the credit risk of the underlying reference security.

These three types of credit derivatives have had limited success compared with the fourth type of credit derivative, the **credit default swap (CDS)**. The credit default swap, in particular, has achieved much success by capturing many of the essential features of insurance while avoiding the high degree of consumer regulations that are typically associated with traditional insurance products.

In a CDS, one party—the credit protection buyer, who is seeking credit protection against a third party—makes a series of regularly scheduled payments to the other party, the credit protection seller. The seller makes no payments until a credit event occurs. A declaration of bankruptcy is clearly a credit event, but there are other types of credit events, such as a failure to make a scheduled payment or an involuntary restructuring. The CDS contract specifies what constitutes a credit event, and the industry has a procedure for declaring credit events, though that does not guarantee the parties will not end up in court arguing over whether something was or was not a credit event.

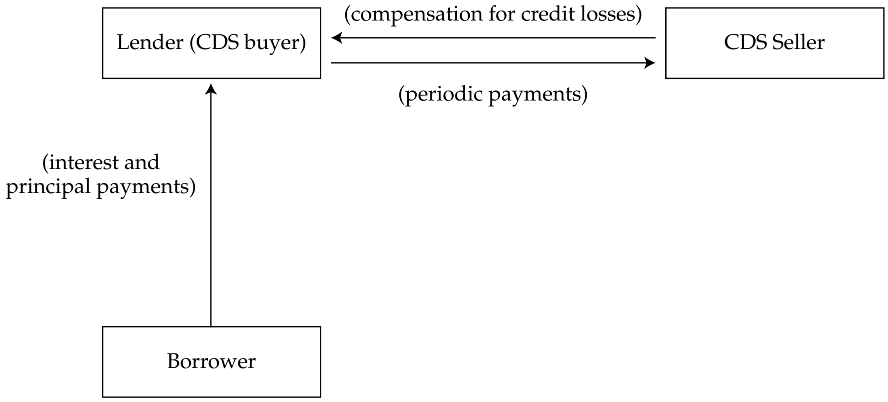
Formally, a credit default swap is defined as follows:

A credit default swap is a derivative contract between two parties, a credit protection buyer and a credit protection seller, in which the buyer makes a series of cash payments to the seller and receives a promise of compensation for credit losses resulting from the default of a third party.

A CDS is conceptually a form of insurance. Sellers of CDSs, oftentimes banks or insurance companies, collect periodic payments and are required to pay out if a loss occurs from the default of a third party. These payouts could take the form of restitution of the defaulted amount or the party holding the defaulting asset could turn it over to the CDS seller and receive a fixed amount. The most common approach is for the payout to be determined by an auction to estimate the market value of the defaulting debt. Thus, CDSs effectively provide coverage against a loss in return for the protection buyer paying a premium to the protection seller, thereby taking the form of insurance against credit loss. Although insurance contracts have certain legal characteristics that are not found in credit default swaps, the two instruments serve similar purposes and operate in virtually the same way: payments made by one party in return for a promise to cover losses incurred by the other.

Exhibit 5 illustrates the typical use of a CDS by a lender. The lender is exposed to the risk of non-payment of principal and interest. The lender lays off this risk by purchasing a CDS from a CDS seller. The lender—now the CDS buyer—promises to make a series of periodic payments to the CDS seller, who then stands ready to compensate the CDS buyer for credit losses.

Exhibit 5 Using a Credit Default Swap to Hedge the Credit Risk of a Loan



Clearly, the CDS seller is betting on the borrower's not defaulting or—more generally, as insurance companies operate—that the total payouts it is responsible for are less than the total payments collected. Of course, most insurance companies are able to do this by having reliable actuarial statistics, diversifying their risk, and selling some of the risk to other insurance companies. Actuarial statistics are typically quite solid. Average claims for life, health, and casualty insurance are well documented, and insurers can normally set premiums to cover losses and operate at a reasonable profit. Although insurance companies try to manage some of their risks at the micro level (e.g., charging smokers more for life and health insurance), most of their risk management is at the macro level, wherein they attempt to make sure their risks are not concentrated. Thus, they avoid selling too much homeowners insurance to individuals in tornado-prone areas. If they have such an exposure, they can use the reinsurance market to sell some of the risk to other companies that are not overexposed to that risk. Insurance companies attempt to diversify their risks and rely on the principle of uncorrelated risks, which plays such an important role in portfolio management. A well-diversified insurance company, like a well-diversified portfolio, should be able to earn a return commensurate with its assumed risk in the long run.

Credit default swaps should operate the same way. Sellers of CDSs should recognize when their credit risk is too concentrated. When that happens, they become buyers of CDSs from other parties or find other ways to lay off the risk. Unfortunately, during the financial crisis that began in 2007, many sellers of CDSs failed to recognize the high correlations among borrowers whose debt they had guaranteed. One well-known CDS seller, AIG, is a large and highly successful traditional insurance company that got into the business of selling CDSs. Many of these CDSs insured against mortgages. With the growth of the subprime mortgage market, many of these CDS-insured mortgages had a substantial amount of credit risk and were often poorly documented. AIG and many other CDS sellers were thus highly exposed to systemic credit contagion, a situation in which defaults in one area of an economy ripple into another, accompanied by bank weaknesses and failures, rapidly falling equity markets, rising credit risk premiums, and a general loss of confidence in the financial system and the economy. These presumably well-diversified risks guaranteed by CDS sellers, operating as though they were insurance companies, ultimately proved to be poorly diversified. Systemic financial risks can spread more rapidly than fire, health, and casualty risks. Virtually no other risks, except those originating from wars or epidemics, spread in the manner of systemic financial risks.

Thus, to understand and appreciate the importance of the CDS market, it is necessary to recognize how that market can fail. The ability to separate and trade risks is a valuable one. Banks can continue to make loans to their customers, thereby satisfying the customers' needs, while laying off the risk elsewhere. In short, parties not wanting to bear certain risks can sell them to parties wanting to assume certain risks. If all parties do their jobs correctly, the markets and the economy work more efficiently. If, as in the case of certain CDS sellers, not everyone does a good job of managing risk, there can be serious repercussions. In the case of AIG and some other companies, taxpayer bailouts were the ultimate price paid to keep these large institutions afloat so that they could continue to provide their other critical services to consumers. The rules proposed in the new OTC derivatives market regulations—which call for greater regulation and transparency of OTC derivatives and, in particular, CDSs—have important implications for the future of this market and these instruments.

EXAMPLE 6

Options and Credit Derivatives

1 An option provides which of the following?
   - A Either the right to buy or the right to sell an underlying
   - B The right to buy and sell, with the choice made at expiration
   - C The obligation to buy or sell, which can be converted into the right to buy or sell

2 Which of the following is **not** a characteristic of a call option on a stock?
   - A A guarantee that the stock will increase
   - B A specified date on which the right to buy expires
   - C A fixed price at which the call holder can buy the stock

3 A credit derivative is which of the following?
   - A A derivative in which the premium is obtained on credit

B A derivative in which the payoff is borrowed by the seller

C A derivative in which the seller provides protection to the buyer against credit loss from a third party

Solution to 1:

A is correct. An option is strictly the right to buy (a call) or the right to sell (a put). It does not provide both choices or the right to convert an obligation into a right.

Solution to 2:

A is correct. A call option on a stock provides no guarantee of any change in the stock price. It has an expiration date, and it provides for a fixed price at which the holder can exercise the option, thereby purchasing the stock.

Solution to 3:

C is correct. Credit derivatives provide a guarantee against loss caused by a third party's default. They do not involve borrowing the premium or the payoff.

8 TYPES OF DERIVATIVES: ASSET-BACKED SECURITIES AND HYBRIDS

b contrast forward commitments with contingent claims;

c define forward contracts, futures contracts, options (calls and puts), swaps, and credit derivatives and compare their basic characteristics;

Although these instruments are covered in more detail in the fixed-income material, we would be remiss if we failed to include them with derivatives. But we will give them only light coverage here.

As discussed earlier, derivatives take (derive) their value from the value of the underlying, as do mutual funds and exchange-traded funds (ETFs). A mutual fund or an ETF holding bonds is virtually identical to the investor holding the bonds directly. Asset-backed securities (ABSs) take this concept a step further by altering the payment streams. ABSs typically divide the payments into slices, called tranches, in which the priority of claims has been changed from equivalent to preferential. For example, in a bond mutual fund or an ETF, all investors in the fund have equal claims, and so the rate of return earned by each investor is exactly the same. If a portfolio of the same bonds were assembled into an ABS, some investors in the ABS would have claims that would supersede those of other investors. The differential nature of these claims becomes relevant when either prepayments or defaults occur.

Prepayments mostly affect only mortgages. When a portfolio of mortgages is assembled into an ABS, the resulting instrument is called a **collateralized mortgage obligation** (CMO). Commonly but not always, the credit risk has been reduced or eliminated, perhaps by a CDS, as discussed earlier. When homeowners pay off their mortgages early due to refinancing at lower rates, the holders of the mortgages suffer losses. They expected to receive a stream of returns that is now terminated. The funds that were previously earning a particular rate will now have to be invested to earn a lower rate. These losses are the mirror images of the gains homeowners make when they proudly proclaim that they refinanced their mortgages and substantially lowered their payments.

CMOs partition the claims against these mortgages into different tranches, which are typically called A, B, and C. Class C tranches bear the first wave of prepayments until that tranche has been completely repaid its full principal investment. At that point, the Class B tranche holders bear the next prepayments until they have been fully repaid. The Class A tranche holders then bear the next wave of prepayments.[11] Thus, the risk faced by the various tranche holders is different from that of a mutual fund or ETF, which would pass the returns directly through such that investors would all receive the same rates of return. Therefore, the expected returns of CMO tranches vary and are commensurate with the prepayment risk they assume. Some CMOs are also characterized by credit risk, perhaps a substantial amount, from subprime mortgages.

When bonds or loans are assembled into ABSs, they are typically called **collateralized bond obligations** (CBOs) or **collateralized loan obligations** (CLOs). These instruments (known collectively as **collateralized debt obligations**, or CDOs) do not traditionally have much prepayment risk but they do have credit risk and oftentimes a great deal of it. The CDO structure allocates this risk to tranches that are called senior, mezzanine, or junior tranches (the last sometimes called equity tranches). When defaults occur, the junior tranches bear the risk first, followed by the mezzanine tranches, and then the senior tranches. The expected returns of the tranches vary according to the perceived credit risk, with the senior tranches having the highest credit quality and the junior the lowest. Thus, the senior tranches have the lowest expected returns and the junior tranches have the highest.

An asset-backed security is formally defined as follows:

> *An asset-backed security is a derivative contract in which a portfolio of debt instruments is assembled and claims are issued on the portfolio in the form of tranches, which have different priorities of claims on the payments made by the debt securities such that prepayments or credit losses are allocated to the most-junior tranches first and the most-senior tranches last.*

ABSs seem to have only an indirect and subtle resemblance to options, but they are indeed options. They promise to make a series of returns that are typically steady. These returns can be lowered if prepayments or defaults occur. Thus, they are contingent on prepayments and defaults. Take a look again at Exhibit 4, Panel B (the profit and payoff of a short put option). If all goes well, there is a fixed return. If something goes badly, the return can be lowered, and the worse the outcome, the lower the return. Thus, holders of ABSs have effectively written put options.

This completes the discussion of contingent claims. Having now covered forward commitments and contingent claims, the final category of derivative instruments is more or less just a catch-all category in case something was missed.

8.1 Hybrids

The instruments just covered encompass all the fundamental instruments that exist in the derivatives world. Yet, the derivatives world is truly much larger than implied by what has been covered here. We have not covered and will touch only lightly on the many hybrid instruments that combine derivatives, fixed-income securities, currencies, equities, and commodities. For example, options can be combined with bonds to form either callable bonds or convertible bonds. Swaps can be combined with options to form swap payments that have upper and lower limits. Options can be combined

11 The reference to only three tranches is just a general statement. There are many more types of tranches. Our discussion of the three classes is for illustrative purposes only and serves to emphasize that there are high-priority claims, low-priority claims, and other claims somewhere in the middle.

with futures to obtain options on futures. Options can be created with swaps as the underlying to form swaptions. Some of these instruments will be covered later. For now, you should just recognize that the possibilities are almost endless.

We will not address these hybrids directly, but some are covered elsewhere in the curriculum. The purpose of discussing them here is for you to realize that derivatives create possibilities not otherwise available in their absence. This point will lead to a better understanding of why derivatives exist, a topic we will get to very shortly.

EXAMPLE 7

Forward Commitments versus Contingent Claims

1 Which of the following is **not** a forward commitment?
   - A An agreement to take out a loan at a future date at a specific rate
   - B An offer of employment that must be accepted or rejected in two weeks
   - C An agreement to lease a piece of machinery for one year with a series of fixed monthly payments

2 Which of the following statements is true about contingent claims?
   - A Either party can default to the other.
   - B The payoffs are linearly related to the performance of the underlying.
   - C The most the long can lose is the amount paid for the contingent claim.

Solution to 1:

B is correct. Both A and C are commitments to engage in transactions at future dates. In fact, C is like a swap because the party agrees to make a series of future payments and in return receives temporary use of an asset whose value could vary. B is a contingent claim. The party receiving the employment offer can accept it or reject it if there is a better alternative.

Solution to 2:

C is correct. The maximum loss to the long is the premium. The payoffs of contingent claims are not linearly related to the underlying, and only one party, the short, can default.

9 DERIVATIVES UNDERLYINGS

b contrast forward commitments with contingent claims;

c define forward contracts, futures contracts, options (calls and puts), swaps, and credit derivatives and compare their basic characteristics;

Before discussing the purposes and benefits of derivatives, we need to clarify some points that have been implied so far. We have alluded to certain underlying assets, this section will briefly discuss the underlyings more directly.

9.1 Equities

Equities are one of the most popular categories of underlyings on which derivatives are created. There are two types of equities on which derivatives exist: individual stocks and stock indexes. Derivatives on individual stocks are primarily options. Forwards, futures, and swaps on individual stocks are not widely used. Index derivatives in the form of options, forwards, futures, and swaps are very popular. Index swaps, more often called equity swaps, are quite popular and permit investors to pay the return on one stock index and receive the return on another index or a fixed rate. They can be very useful in asset allocation strategies by allowing an equity manager to increase or reduce exposure to an equity market or sector without trading the individual securities.

In addition, options on stocks are frequently used by companies as compensation and incentives for their executives and employees. These options are granted to provide incentives to work toward driving the stock price up and can result in companies paying lower cash compensation. Some companies also issue warrants, which are options sold to the public that allow the holders to exercise them and buy shares directly from the companies.

9.2 Fixed-Income Instruments and Interest Rates

Options, forwards, futures, and swaps on bonds are widely used. The problem with creating derivatives on bonds, however, is that there are almost always many issues of bonds. A single issuer, whether it is a government or a private borrower, often has more than one bond issue outstanding. For futures contracts, with their standardization requirements, this problem is particularly challenging. What does it mean to say that a futures contract is on a German bund, a US Treasury note, or a UK gilt? The most common solution to this problem is to allow multiple issues to be delivered on a single futures contract. This feature adds some interesting twists to the pricing and trading strategies of these instruments.

Until now, we have referred to the underlying as an *asset*. Yet, one of the largest derivative underlyings is not an asset. It is simply an interest rate. An interest rate is not an asset. One cannot hold an interest rate or place it on a balance sheet as an asset. Although one can hold an instrument that pays an interest rate, the rate itself is not an asset. But there are derivatives in which the rate, not the instrument that pays the rate, is the underlying. In fact, we have already covered one of these derivatives: The plain vanilla interest rate swap in which Libor is the underlying.[12] Instead of a swap, an interest rate derivative could be an option. For example, a call option on 90-day Libor with a strike of 5% would pay off if at expiration Libor exceeds 5%. If Libor is below 5%, the option simply expires unexercised.

Interest rate derivatives are the most widely used derivatives. With that in mind, we will be careful in using the expression *underlying asset* and will use the more generic *underlying.*

9.3 Currencies

Currency risk is a major factor in global financial markets, and the currency derivatives market is extremely large. Options, forwards, futures, and swaps are widely used. Currency derivatives can be complex, sometimes combining elements of other underlyings. For example, a currency swap involves two parties making a series of interest rate payments to each other in different currencies. Because interest rates

12 As you will see later, there are also futures in which the underlying is an interest rate (Eurodollar futures) and forwards in which the underlying is an interest rate (forward rate agreements, or FRAs).

and currencies are both subject to change, a currency swap has two sources of risk. Although this instrument may sound extremely complicated, it merely reflects the fact that companies operating across borders are subject to both interest rate risk and currency risk and currency swaps are commonly used to manage those risks.

9.4 Commodities

Commodities are resources, such as food, oil, and metals, that humans use to sustain life and support economic activity. Because of the economic principle of comparative advantage, countries often specialize in the production of certain resources. Thus, the commodities market is extremely large and subject to an almost unimaginable array of risks. One need only observe how the price of oil moves up as tension builds in the Middle East or how the price of orange juice rises on a forecast of cold weather in Florida.

Commodity derivatives are widely used to speculate in and manage the risk associated with commodity price movements. The primary commodity derivatives are futures, but forwards, swaps, and options are also used. The reason that futures are in the lead in the world of commodities is simply history. The first futures markets were futures on commodities. The first futures exchange, the Chicago Board of Trade, was created in 1848, and until the creation of currency futures in 1972, there were no futures on any underlying except commodities.

There has been a tendency to think of the commodities world as somewhat separate from the financial world. Commodity traders and financial traders were quite different groups. Since the creation of financial futures, however, commodity and financial traders have become relatively homogeneous. Moreover, commodities are increasingly viewed as an important asset class that should be included in investment strategies because of their ability to help diversify portfolios.

9.5 Credit

As we previously discussed, credit is another underlying and quite obviously not an asset. Credit default swaps (CDSs) and collateralized debt obligations (CDOs) were discussed extensively in an earlier section. These instruments have clearly established that credit is a distinct underlying that has widespread interest from a trading and risk management perspective. In addition, to the credit of a single entity, credit derivatives are created on multiple entities. CDOs themselves are credit derivatives on portfolios of credit risks. In recent years, indexes of CDOs have been created, and instruments based on the payoffs of these CDO indexes are widely traded.

9.6 Other

This category is included here to capture some of the really unusual underlyings. One in particular is weather. Although weather is hardly an asset, it is certainly a major force in how some entities perform. For example, a ski resort needs snow, farmers need an adequate but not excessive amount of rain, and public utilities experience strains on their capacity during temperature extremes. Derivatives exist in which the payoffs are measured as snowfall, rainfall, and temperature. Although these derivatives have not been widely used—because of some complexities in pricing, among other things—they continue to exist and may still have a future. In addition, there are derivatives on electricity, which is also not an asset. It cannot be held in the traditional sense because it is created and consumed almost instantaneously. Another unusual type of derivative is based on disasters in the form of insurance claims.

Financial institutions will continue to create derivatives on all types of risks and exposures. Most of these derivatives will fail because of little trading interest, but a few will succeed. If that speaks badly of derivatives, it must be remembered that most small businesses fail, most creative ideas fail, and most people who try to become professional entertainers or athletes fail. It is the sign of a healthy and competitive system that only the very best survive.

THE SIZE OF THE DERIVATIVES MARKET

In case anyone thinks that the derivatives market is not large enough to justify studying, we should consider how big the market is. Unfortunately, gauging the size of the derivatives market is not a simple task. OTC derivatives contracts are private transactions. No reporting agency gathers data, and market size is not measured in traditional volume-based metrics, such as shares traded in the stock market. Complicating things further is the fact that derivatives underlyings include equities, fixed-income securities, interest rates, currencies, commodities, and a variety of other underlyings. All these underlyings have their own units of measurement. Hence, measuring how "big" the underlying derivatives markets are is like trying to measure how much fruit consumers purchase; the proverbial mixing of apples, oranges, bananas, and all other fruits.

The exchange-listed derivatives market reports its size in terms of volume, meaning the number of contracts traded. Exchange-listed volume, however, is an inconsistent number. For example, US Treasury bond futures contracts trade in units covering $100,000 face value. Eurodollar futures contracts trade in units covering $1,000,000 face value. Crude oil trades in 1,000-barrel (42 gallons each) units. Yet, one traded contract of each gets equal weighting in volume totals.

FIA (a global trade organization for futures, options, and centrally cleared derivatives markets) publishes detailed information about the industry. For 2017, global trading volume was 14.8 billion futures contracts, 10.4 billion options contracts, and a combined total of 25.2 billion contracts. Of nine futures instrument types, the largest three categories were interest rate futures (21.4% of global trading), equity index futures (16.9%), and currency futures (14.6%). Of nine options instrument types, the largest three categories were equity index options (48.4% of global trading), individual equity options (33.5%), and currency options (7.9%). Because options and futures contracts are typically short lived, the number of open-interest contracts is substantially less than the amount of trading volume. For the end of December 2017, open-interest futures contracts were 0.24 billion, open-interest option contracts were 0.60 billion, and the combined total was 0.84 billion contracts.

OTC volume is even more difficult to measure. There is no count of the number of contracts that trade. In fact, *volume* is an almost meaningless concept in OTC markets because any notion of volume requires a standardized size. If a customer goes to a swaps dealer and enters into a swap to hedge a $50 million loan, there is no measure of how much volume that transaction generated. The $50 million swap's notional principal, however, does provide a measure to some extent. Forwards, swaps, and OTC options all have notional principals, so they can be measured in that manner. Another measure of the size of the derivatives market is the market value of these contracts. As noted, forwards and swaps start with zero market value, but their market value changes as market conditions change. Options do not start with zero market value and almost always have a positive market value until expiration, when some options expire out of the money.

The OTC industry has taken both of these concepts—notional principal and market value—as measures of the size of the market. Notional principal is probably a more accurate measure. The amount of a contract's notional principal is unambiguous: It is written into the contract and the two parties cannot disagree over it. Yet, notional principal terribly overstates the amount of money actually at risk. For example, a $50 million notional principal swap will have nowhere near $50 million at risk. The payments on such a swap are merely the net of two opposite series of interest payments on $50 million. The market value of such a swap is the present value of one stream of payments minus the present value of the other. This market value figure will always be well below the notional principal. Thus, market value seems like a better measure except that, unlike notional principal, it is not unambiguous. Market value requires measurement, and two parties can disagree on the market value of the same transaction.

Notional principal and market value estimates for the global OTC derivatives market are collected semi-annually by the Bank for International Settlements of Basel, Switzerland, and published on its website (http://www.bis.org/statistics/derstats.htm). At the end of 2017, notional principal was more than $532 trillion and market value was about $27 trillion. A figure of $600 trillion is an almost unfathomable number and, as noted, is a misleading measure of the amount of money at risk.[13] The market value figure of $11 trillion is a much more realistic measure, but as noted, it is less accurate, relying on estimates provided by banks. Interest rate contracts constituted 81.9% of the total notional amount outstanding. The relative sizes of the other categories were 16.4% for foreign exchange contracts and less than 2% for credit derivatives, equity-linked contracts, and commodity contracts.

The exchange-listed and OTC markets use different measures and each of those measures is subject to severe limitations. About all we can truly say for sure about the derivatives market is, "It is big."

10 THE PURPOSES AND BENEFITS OF DERIVATIVES

e describe purposes of, and controversies related to, derivative markets;

Economic historians know that derivatives markets have existed since at least the Middle Ages. It is unclear whether derivatives originated in the Asian rice markets or possibly in medieval trade fairs in Europe. We do know that the origin of modern futures markets is the creation of the Chicago Board of Trade in 1848. To understand why derivatives markets exist, it is useful to take a brief look at why the Chicago Board of Trade was formed.

In the middle of the 19th century, midwestern America was rapidly becoming the center of agricultural production in the United States. At the same time, Chicago was evolving into a major American city, a hub of transportation and commerce. Grain markets in Chicago were the central location to which midwestern farmers brought their wheat, corn, and soybeans to sell. Unfortunately, most of these products arrived at approximately the same time of the year, September through November. The storage facilities in Chicago were strained beyond capacity. As a result, prices would fall tremendously and some farmers reportedly found it more economical to dump their grains in the Chicago River rather than transport them back to the farm. At other times of the year, prices would rise steeply. A group of businessmen saw this situation as unnecessary volatility and a waste of valuable produce. To deal with this problem, they created the Chicago Board of Trade and a financial instrument called the "to-arrive" contract. A farmer could sell a to-arrive contract at any time during the year. This contract fixed the price of the farmer's grain on the basis of delivery in Chicago at a specified later date. Grain is highly storable, so farmers can hold on to the grain and deliver it at almost any later time. This plan substantially reduced seasonal market volatility and made the markets work much better for all parties.

The traders in Chicago began to trade these contracts, speculating on movements in grain prices. Soon, it became apparent that an important and fascinating market had developed. Widespread hedging and speculative interest resulted in substantial

13 To put it in perspective, it would take almost 17 million years for a clock to tick off 532 trillion seconds!

market growth, and about 80 years later, a clearinghouse and a performance guarantee were added, thus completing the evolution of the to-arrive contract into today's modern futures contract.

Many commodities and all financial assets that underlie derivatives contracts are not seasonally produced. Hence, this initial motivation for futures markets is only a minor advantage of derivatives markets today. But there are many reasons why derivative markets serve an important and useful purpose in contemporary finance.

10.1 Risk Allocation, Transfer, and Management

Until the advent of derivatives markets, risk management was quite cumbersome. Setting the actual level of risk to the desired level of risk required engaging in transactions in the underlyings. Such transactions typically had high transaction costs and were disruptive of portfolios. In many cases, it is quite difficult to fine-tune the level of risk to the desired level. From the perspective of a risk taker, it was quite costly to buy risk because a large amount of capital would be required.

Derivatives solve these problems in a very effective way: They allow trading the risk without trading the instrument itself. For example, consider a stockholder who wants to reduce exposure to a stock. In the pre-derivatives era, the only way to do so was to sell the stock. Now, the stockholder can sell futures, forwards, calls, or swaps, or buy put options, all while retaining the stock. For a company founder, these types of strategies can be particularly useful because the founder can retain ownership and probably board membership. Many other excellent examples of the use of derivatives to transfer risk are covered elsewhere in the curriculum. The objective at this point is to establish that derivatives provide an effective method of transferring risk from parties who do not want the risk to parties who do. In this sense, risk allocation is improved within markets and, indeed, the entire global economy.

The overall purpose of derivatives is to obtain more effective risk management within companies and the entire economy. Although some argue that derivatives do not serve this purpose very well (we will discuss this point in Section 11), for now you should understand that derivatives can improve the allocation of risk and facilitate more effective risk management for both companies and economies.

10.2 Information Discovery

One of the advantages of futures markets has been described as *price discovery*. A futures price has been characterized by some experts as a revelation of some information about the future. Thus, a futures price is sometimes thought of as predictive. This statement is not strictly correct because futures prices are not really forecasts of future spot prices. They provide only a little more information than do spot prices, but they do so in a very efficient manner. The markets for some underlyings are highly decentralized and not very efficient. For example, what is gold worth? It trades in markets around the world, but probably the best place to look is at the gold futures contract expiring soonest. What is the value of the S&P 500 Index when the US markets are not open? As it turns out, US futures markets open before the US stock market opens. The S&P 500 futures price is frequently viewed as an indication of where the stock market will open.

Derivative markets can, however, convey information not impounded in spot markets. By virtue of the fact that derivative markets require less capital, information can flow into the derivative markets before it gets into the spot market. The difference may well be only a matter of minutes or possibly seconds, but it can provide the edge to astute traders.

Finally, we should note that futures markets convey another simple piece of information: What price would one accept to avoid uncertainty? If you hold a stock worth $40 and could hedge the next 12 months' uncertainty, what locked-in price should you expect to earn? As it turns out, it should be the price that guarantees the risk-free rate minus whatever dividends would be paid on the stock. Derivatives—specifically, futures, forwards, and swaps—reveal the price that the holder of an asset could take and avoid the risk.

What we have said until now applies to futures, forwards, and swaps. What about options? As you will learn later, given the underlying and the type of option (call or put), an option price reflects two characteristics of the option (exercise price and time to expiration), three characteristics of the underlying (price, volatility, and cash flows it might pay), and one general macroeconomic factor (risk-free rate). Only one of these factors, volatility, is not relatively easy to identify. But with the available models to price the option, we can infer what volatility people are using from the actual market prices at which they execute trades. That volatility, called **implied volatility**, measures the expected risk of the underlying. It reflects the volatility that investors use to determine the market price of the option. Knowing the expected risk of the underlying asset is an extremely useful piece of information. In fact, for options on broad-based market indexes, such as the S&P 500, the implied volatility is a good measure of the general level of uncertainty in the market. Some experts have even called it a measure of fear. Thus, options provide information about what investors think of the uncertainty in the market, if not their fear of it.[14]

In addition, options allow the creation of trading strategies that cannot be done by using the underlying. As the exhibits on options explained, these strategies provide asymmetrical performance: limited movement in one direction and movement in the other direction that changes with movements in the underlying.

10.3 Operational Advantages

We noted earlier that derivatives have lower transaction costs than the underlying. The transaction costs of derivatives can be high relative to the value of the derivatives, but these costs are typically low relative to the value of the underlying. Thus, an investor who wants to take a position in, say, an equity market index would likely find it less costly to use the futures to get a given degree of exposure than to invest directly in the index to get that same exposure.

Derivative markets also typically have greater liquidity than the underlying spot markets, a result of the smaller amount of capital required to trade derivatives than to get the equivalent exposure directly in the underlying. Futures margin requirements and option premiums are quite low relative to the cost of the underlying.

One other extremely valuable operational advantage of derivative markets is the ease with which one can go short. With derivatives, it is nearly as easy to take a short position as to take a long position, whereas for the underlying asset, it is almost always much more difficult to go short than to go long. In fact, for many commodities, short selling is nearly impossible.

14 The Chicago Board Options Exchange publishes a measure of the implied volatility of the S&P 500 Index option, which is called the VIX (volatility index). The VIX is widely followed and is cited as a measure of investor uncertainty and sometimes fear.

10.4 Market Efficiency

In the study of portfolio management, you learn that an efficient market is one in which no single investor can consistently earn returns in the long run in excess of those commensurate with the risk assumed. Of course, endless debates occur over whether equity markets are efficient. No need to resurrect that issue here, but let us proceed with the assumption that equity markets—and, in fact, most free and competitive financial markets—are reasonably efficient. This assumption does not mean that abnormal returns can never be earned, and indeed prices do get out of line with fundamental values. But competition, the relatively free flow of information, and ease of trading tend to bring prices back in line with fundamental values. Derivatives can make this process work even more rapidly.

When prices deviate from fundamental values, derivative markets offer less costly ways to exploit the mispricing. As noted earlier, less capital is required, transaction costs are lower, and short selling is easier. We also noted that as a result of these features, it is possible, indeed likely, that fundamental value will be reflected in the derivatives markets before it is restored in the underlying market. Although this time difference could be only a matter of minutes, for a trader seeking abnormal returns, a few minutes can be a valuable opportunity.

All these advantages of derivatives markets make the financial markets in general function more effectively. Investors are far more willing to trade if they can more easily manage their risk, trade at lower cost and with less capital, and go short more easily. This increased willingness to trade increases the number of market participants, which makes the market more liquid. A very liquid market may not automatically be an efficient market, but it certainly has a better chance of being one.

Even if one does not accept the concept that financial markets are efficient, it is difficult to say that markets are not more effective and competitive with derivatives. Yet, many blame derivatives for problems in the market. Let us take a look at these arguments.

11 CRITICISMS AND MISUSES OF DERIVATIVES

e describe purposes of, and controversies related to, derivative markets;

The history of financial markets is filled with extreme ups and downs, which are often called bubbles and crashes. Bubbles occur when prices rise for a long time and appear to exceed fundamental values. Crashes occur when prices fall rapidly. Although bubbles, if they truly exist, are troublesome, crashes are even more so because nearly everyone loses substantial wealth in a crash. A crash is then typically followed by a government study commissioned to find the causes of the crash. In the last 30 years, almost all such studies have implicated derivatives as having some role in causing the crash. Of course, because derivatives are widely used and involve a high degree of leverage, it is a given that they would be seen in a crash. It is unclear whether derivatives are the real culprit or just the proverbial smoking gun used by someone to do something wrong.

The two principal arguments against derivatives are that they are such speculative devices that they effectively permit legalized gambling and that they destabilize the financial system. Let us look at these points more closely.

11.1 Speculation and Gambling

As noted earlier, derivatives are frequently used to manage risk. In many contexts, this use involves hedging or laying off risk. Naturally, for hedging to work, there must be speculators. Someone must accept the risk. Derivatives markets are unquestionably attractive to speculators. All the benefits of derivatives draw speculators in large numbers, and indeed they should. The more speculators that participate in the market, the cheaper it is for hedgers to lay off risk. These speculators take the form of hedge funds and other professional traders who willingly accept risk that others need to shed. In recent years, the rapid growth of these types of investors has been alarming to some but almost surely has been beneficial for all investors.

Unfortunately, the general image of speculators is not a good one. Speculators are often thought to be short-term traders who attempt to exploit temporary inefficiencies, caring little about long-term fundamental values. The profits from short-term trading are almost always taxed more heavily than the profits from long-term trading, clearly targeting and in some sense punishing speculators. Speculators are thought to engage in price manipulation and to trade at extreme prices.[15] All of this type of trading is viewed more or less as just a form of gambling.

Yet, there are notable differences between gambling and speculation. Gambling typically benefits only a limited number of participants and does not generally help society as a whole. But derivatives trading brings extensive benefits to financial markets, as explained earlier, and thus does benefit society as a whole. In short, the benefits of derivatives are broad, whereas the benefits of gambling are narrow.

Nonetheless, the argument that derivatives are a form of legalized gambling will continue to be made. Speculation and gambling are certainly both forms of financial risk taking, so these arguments are not completely off base. But insurance companies speculate on loss claims, mutual funds that invest in stocks speculate on the performance of companies, and entrepreneurs go up against tremendous odds to speculate on their own ability to create successful businesses. These so-called speculators are rarely criticized for engaging in a form of legalized gambling, and indeed entrepreneurs are praised as the backbone of the economy. Really, all investment is speculative. So, why is speculation viewed as such a bad thing by so many? The answer is unclear.

11.2 Destabilization and Systemic Risk

The arguments against speculation through derivatives often go a step further, claiming that it is not merely speculation or gambling per se but rather that it has destabilizing consequences. Opponents of derivatives claim that the very benefits of derivatives (low cost, low capital requirements, ease of going short) result in an excessive amount of speculative trading that brings instability to the market. They argue that speculators use large amounts of leverage, thereby subjecting themselves and their creditors to substantial risk if markets do not move in their hoped-for direction. Defaults by speculators can then lead to defaults by their creditors, their creditors' creditors, and so on. These effects can, therefore, be systemic and reflect an epidemic contagion whereby instability can spread throughout markets and an economy, if not the entire world. Given that governments often end up bailing out some banks and insurance companies, society has expressed concern that the risk managed with derivatives must be controlled.

---

15 Politicians and regulators have been especially critical of energy market speculators. Politicians, in particular, almost always blame rising oil prices on speculators, although credit is conspicuously absent for falling oil prices.

This argument is not without merit. Such effects occurred in the Long-Term Capital Management fiasco of 1998 and again in the financial crisis of 2008, in which derivatives, particularly credit default swaps, were widely used by many of the problem entities. Responses to such events typically take the course of calling for more rules and regulations restricting the use of derivatives, requiring more collateral and credit mitigation measures, backing up banks with more capital, and encouraging, if not requiring, OTC derivatives to be centrally cleared like exchange-traded derivatives.

In response, however, we should note that financial crises—including the South Sea and Mississippi bubbles and the stock market crash of 1929, as well as a handful of economic calamities of the 19th and 20th centuries—have existed since the dawn of capitalism. Some of these events preceded the era of modern derivatives markets, and others were completely unrelated to the use of derivatives. Some organizations, such as Orange County, California, in 1994–1995, have proved that derivatives are not required to take on excessive leverage and nearly bring the entity to ruin. Proponents of derivatives argue that derivatives are but one of many mechanisms through which excessive risk can be taken. Derivatives may seem dangerous, and they can be if misused, but there are many ways to take on leverage that look far less harmful but can be just as risky.

Another criticism of derivatives is simply their complexity. Many derivatives are extremely complex and require a high-level understanding of mathematics. The financial industry employs many mathematicians, physicists, and computer scientists. This single fact has made many distrust derivatives and the people who work on them. It is unclear why this reason has tarnished the reputation of the derivatives industry. Scientists work on complex problems in medicine and engineering without public distrust. One explanation probably lies in the fact that scientists create models of markets by using scientific principles that often fail. To a physicist modeling the movements of celestial bodies, the science is reliable and the physicist is unlikely to misapply the science. The same science applied to financial markets is far less reliable. Financial markets are driven by the actions of people who are not as consistent as the movements of celestial bodies. When financial models fail to work as they should, the scientists are often blamed for either building models that are too complex and unable to accurately capture financial reality or misusing those models, such as using poor estimates of inputs. And derivatives, being so widely used and heavily leveraged, are frequently in the center of it all.

EXAMPLE 8

Purposes and Controversies of Derivative Markets

1. Which of the following is **not** an advantage of derivative markets?
   - **A** They are less volatile than spot markets.
   - **B** They facilitate the allocation of risk in the market.
   - **C** They incur lower transaction costs than spot markets.
2. Which of the following pieces of information is **not** conveyed by at least one type of derivative?
   - **A** The volatility of the underlying
   - **B** The most widely used strategy of the underlying
   - **C** The price at which uncertainty in the underlying can be eliminated
3. Which of the following responds to the criticism that derivatives can be destabilizing to the underlying market?

A Market crashes and panics have occurred since long before derivatives existed.

B Derivatives are sufficiently regulated that they cannot destabilize the spot market.

C The transaction costs of derivatives are high enough to keep their use at a minimum level.

Solution to 1:

A is correct. Derivative markets are not by nature more or less volatile than spot markets. They facilitate risk allocation by making it easier and less costly to transfer risk, and their transaction costs are lower than those of spot markets.

Solution to 2:

B is correct. Options do convey the volatility of the underlying, and futures, forwards, and swaps convey the price at which uncertainty in the underlying can be eliminated. Derivatives do not convey any information about the use of the underlying in strategies.

Solution to 3:

A is correct. Derivatives regulation is not more and is arguably less than spot market regulation, and the transaction costs of derivatives are not a deterrent to their use; in fact, derivatives are widely used. Market crashes and panics have a very long history, much longer than that of derivatives.

An important element of understanding and using derivatives is having a healthy respect for their power. Every day, we use chemicals, electricity, and fire without thinking about their dangers. We consume water and drive automobiles, both of which are statistically quite dangerous. Perhaps these risks are underappreciated, but it is more likely the case that most adults learn how to safely use chemicals, electricity, fire, water, and automobiles. Of course, there are exceptions, many of which are foolish, and foolishness is no stranger to the derivatives industry. The lesson here is that derivatives can make our financial lives better, but like chemicals, electricity, and all the rest, we need to know how to use them safely, which is why they are an important part of the CFA curriculum.

Later in the curriculum, you will learn a great deal about how derivatives are priced. At this point, we introduce the pricing of derivatives. This material not only paves the way for a deeper understanding of derivatives but also complements earlier material by helping you understand how derivatives work.

12 ELEMENTARY PRINCIPLES OF DERIVATIVE PRICING

f explain arbitrage and the role it plays in determining prices and promoting market efficiency.

Pricing and valuation are fundamental elements of the CFA Program. The study of fixed-income and equity securities, as well as their application in portfolio management, is solidly grounded on the principle of valuation. In valuation, the question is simple: What is something worth? Without an answer to that question, one can hardly proceed to use that *something* wisely.

Determining what a derivative is worth is similar to determining what an asset is worth. As you learn in the fixed-income and equity readings, value is the present value of future cash flows, with discounting done at a rate that reflects both the opportunity cost of money and the risk. Derivatives valuation applies that same principle but in a somewhat different way.

Think of a derivative as *attached* to an underlying. We know that the derivative *derives* its value from the value of the underlying. If the underlying's value changes, so should the value of the derivative. The underlying takes its value from the discounted present value of the expected future cash flows it offers, with discounting done at a rate reflecting the investor's risk tolerance. But if the value of the underlying is embedded in the value of the derivative, it would be double counting to discount the derivative's expected future cash flows at a risky discount rate. That effect has already been incorporated into the value of the underlying, which goes into the value of the derivative.

Derivatives usually take their values from the underlying by constructing a hypothetical combination of the derivatives and the underlyings that eliminates risk. This combination is typically called a **hedge portfolio**. With the risk eliminated, it follows that the hedge portfolio should earn the risk-free rate. A derivative's value is the price of the derivative that forces the hedge portfolio to earn the risk-free rate.

This principle of derivative valuation relies completely on the ability of an investor to hold or store the underlying asset. Let us take a look at what that means.

12.1 Storage

As noted previously, the first derivatives were agricultural commodities. Most of these commodities can be stored (i.e., held) for a period of time. Some extreme cases, such as oil and gold, which are storable for millions of years, are excellent examples of fully storable commodities. Grains, such as wheat and corn, can be stored for long but not infinite periods of time. Some commodities, such as bananas, are storable for relatively short periods of time. In the CFA Program, we are more interested in financial assets. Equities and currencies have perpetual storability, whereas bonds are storable until they mature.

Storage incurs costs. Commodity storage costs can be quite expensive. Imagine storing 1,000 kilograms of gold or a million barrels of oil. Financial assets, however, have relatively low storage costs. Some assets pay returns during storage. Stocks pay dividends and bonds pay interest. The net of payments offered minus storage costs plays a role in the valuation of derivatives.

An example earlier in this reading illustrates this point. Suppose an investor holds a dividend-paying stock and wants to eliminate the uncertainty of its selling price over a future period of time. Suppose further that the investor enters into a forward contract that commits him to deliver the stock at a later date, for which he will receive a fixed price. With uncertainty eliminated, the investor should earn the risk-free rate, but in fact, he does not. He earns more because while holding the stock, he collects dividends. Therefore, he should earn the risk-free rate *minus* the dividend yield, a concept known as the cost of carry, which will be covered in great detail in later readings. The cost of carry *plus* the dividends he earns effectively means that he makes the risk-free rate. Now, no one is claiming that this is a good way to earn the risk-free rate. There are many better ways to do that, but this strategy could be executed. There is one and only one forward price that guarantees that this strategy earns a return of the risk-free rate minus the dividend yield, or the risk-free rate after accounting for the dividends collected. If the forward price at which contracts are created does not equal this price, investors can take advantage of this discrepancy by engaging in arbitrage, which is discussed in the next section.

Forwards, futures, swaps, and options are all priced in this manner. Hence, they rely critically on the ability to store or hold the asset. Some underlyings are not storable. We previously mentioned electricity. It is produced and consumed almost instantaneously. Weather is also not storable. Fresh fish have very limited storability. Although this absence of storability may not be the reason, derivative markets in these types of underlyings have not been particularly successful, whereas those in underlyings that are more easily storable have often been successful.

The opposite of storability is the ability to go short—that is, to borrow the underlying, sell it, and buy it back later. We discussed earlier that short selling of some assets can be difficult. It is not easy to borrow oil or soybeans. There are ways around this constraint, but derivatives valuation is generally much easier when the underlying can be shorted. This point is discussed in more depth later in the curriculum.

12.2 Arbitrage

What we have been describing is the foundation of the principle of **arbitrage**. In well-functioning markets with low transaction costs and a free flow of information, the same asset cannot sell for more than one price. If it did, someone would buy it in the cheaper market and sell it in the more expensive market, earning a riskless profit. The combined actions of all parties doing this would push up the lower price and push down the higher price until they converged. For this reason, arbitrage is often referred to as the **law of one price**. Of course, for arbitrage to be feasible, the ability to purchase and sell short the asset is important.

Obviously, this rule does not apply to all markets. The same consumer good can easily sell for different prices, which is one reason why people spend so much time shopping on the internet. The costs associated with purchasing the good in the cheaper market and selling it in the more expensive market can make the arbitrage not worthwhile. The absence of information on the very fact that different prices exist would also prevent the arbitrage from occurring. Although the internet and various price-comparing websites reduce these frictions and encourage all sellers to offer competitive prices, consumer goods are never likely to be arbitragable.[16]

Financial markets, of course, are a different matter. Information on securities prices around the world is quite accessible and relatively inexpensive. Most financial markets are fairly competitive because dealers, speculators, and brokers attempt to execute trades at the best prices. Arbitrage is considered a dependable rule in the financial markets. Nonetheless, there are people who purport to make a living as arbitrageurs. How could they exist? To figure that out, first consider some examples of arbitrage.

The simplest case of an arbitrage might be for the same stock to sell at different prices in two markets. If the stock were selling at \$52 in one market and \$50 in another, an arbitrageur would buy the stock at \$50 in the one market and sell it at \$52 in the other. This trade would net an immediate \$2 profit at no risk and would not require the commitment of any of the investor's capital. This outcome would be a strong motivation for all arbitrageurs, and their combined actions would force the lower price up and the higher price down until the prices converged.

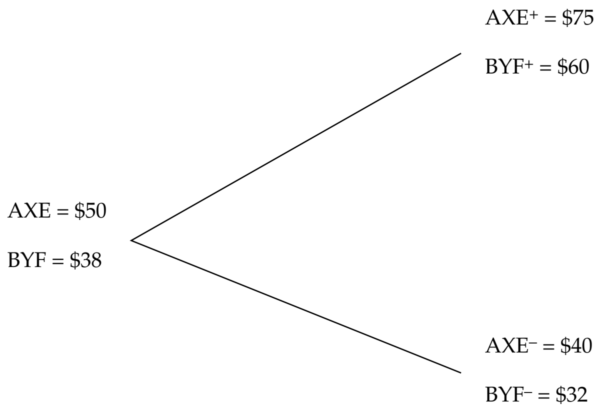
But what would be the final price? It is entirely possible that \$50 is the true fundamental value and \$52 is too high. Or \$52 could be the true fundamental value and \$50 is too low. Or the true fundamental value could lie somewhere between the two. Arbitrage does not tell us the true fundamental value. It is not an *absolute* valuation

16 If the same consumer good sells for different prices in markets with a relatively free flow of information (e.g., via price-comparing websites), it still may not be possible to truly arbitrage. Buying the good at a lower price and selling it at a higher price but less than the price of the most expensive seller may not be practical, but the most expensive seller may be driven out of business. When everyone knows what everyone else is charging, the same effect of arbitrage can still occur.

methodology, such as the discounted cash flow equity valuation model. It is a *relative* valuation methodology. It tells us the correct price of one asset or derivative *relative to* another asset or derivative.

Now, consider another situation, illustrated in Exhibit 6. Observe that we have one stock, AXE Electronics, that today is worth \$50 and one period later will be worth either \$75 or \$40. We will denote these prices as AXE = \$50, AXE^+ = \$75, and AXE^- = \$40. Another stock, BYF Technology, is today worth \$38 and one period later will be worth \$60 or \$32. Thus, BYF = \$38, BYF^+ = \$60, and BYF^- = \$32. Assume that the risk-free borrowing and lending rate is 4%. Also assume no dividends are paid on either stock during the period covered by this example.

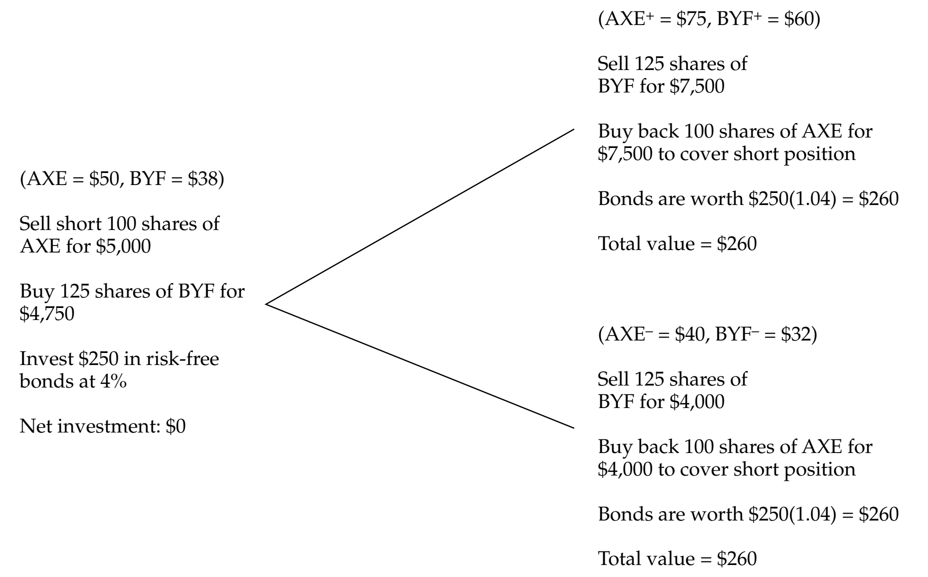
Exhibit 6 Arbitrage Opportunity with Stock AXE, Stock BYF, and a Risk-Free Bond



Note: The risk-free rate is 4%.

The opportunity exists to make a profit at no risk without committing any of our funds, as demonstrated in Exhibit 7. Suppose we borrow 100 shares of stock AXE, which is selling for \$50, and sell them short, thereby receiving \$5,000. We take \$4,750 and purchase 125 shares of stock BYF (125 × \$38 = \$4,750). We invest the remaining \$250 in risk-free bonds at 4%. This transaction will not require us to use any funds of our own: The short sale will be sufficient to fund the investment in BYF and leave money to invest in risk-free bonds.

Exhibit 7 Execution of Arbitrage Transaction with Stock AXE, Stock BYF, and a Risk-Free Bond



If the top outcome in Exhibit 7 occurs, we sell the 125 shares of BYF for 125 × $60 = $7,500. This amount is sufficient to buy back the 100 shares of AXE, which is selling for $75. But we will also have the bonds, which are worth $250 × 1.04 = $260. If the bottom outcome occurs, we sell the 125 shares of BYF for 125 × $32 = $4,000—enough money to buy back the 100 shares of AXE, which is selling for $40. Again, we will have the risk-free bonds, worth $260. Regardless of the outcome, we end up with $260.

Recall that we invested no money of our own and end up with a sure $260. It should be apparent that this transaction is extremely attractive, so everyone would do it. The combined actions of multiple investors would drive down the price of AXE and/or drive up the price of BYF until an equilibrium is reached, at which point this transaction would no longer be profitable. As noted earlier, we cannot be sure of the correct fundamental price, but let us assume that BYF's price remains constant. Then AXE would fall to $47.50. Alternatively, if we assume that AXE's price remains constant, then the price of BYF would rise to $40. These values are obtained by noting that the prices for both outcomes occur according to the ratio 1.25 ($75/$60 = 1.25; $40/$32 = 1.25). Thus, their initial prices should be consistent with that ratio. If BYF is $38, AXE should be $38 × 1.25 = $47.50. If AXE is $50, BYF should be $40.00 because $40.00 × 1.25 = $50. Of course, the two prices could settle in between. Arbitrage is only a relative pricing method. It prices the two stocks in relation to each other but does not price either on the basis of its own fundamentals.

Of course, this example is extremely simplified. Clearly, a stock price can change to more than two other prices. Also, if a given stock is at one price, another stock may be at any other price. We have created a simple case here to illustrate a point. But as you will learn later in the curriculum, when derivatives are involved, the simplification here is relatively safe. As we know, the price of a derivative is determined by the price of the underlying. Hence, when the underlying is at one particular price, the derivative's price will be determined by that price. The two assets need not be two stocks; one can be a stock and the other can be a derivative on the stock.

To see that point, consider another type of arbitrage opportunity that involves a forward contract. Recall from the previous example that at the start, AXE sells for $50. Suppose we borrow $50 at 4% interest by issuing a risk-free bond, use the money to buy one share of stock AXE, and simultaneously enter into a forward contract to sell this share at a price of \$54 one period later. The stock will then move to either \$75 or \$40 in the next period. The forward contract requires that we deliver the stock and accept \$54 for it. And of course, we will owe \$50 × 1.04 = \$52 on the loan.

Now consider the two outcomes. Regardless of the outcome, the end result is the same. The forward contract fixes the delivery price of the stock at \$54:

AXE goes to \$75	
Deliver stock to settle forward contract	+ \$54
Pay back loan	– \$52
Net	+ \$2
AXE goes to \$40	
Deliver stock to settle forward contract	+ \$54
Pay back loan	– \$52
Net	+ \$2

In either case, we made \$2, free and clear. In fact, we can even accommodate the possibility of more than two future prices for AXE and we will always make \$2.[17] The key point is that we faced no risk and did not have to invest any of our own money, but ended up with \$2, which is clearly a good trade. The \$2 is an arbitrage profit. But where did it originate?

It turns out that the forward price, \$54, was an inappropriate price given current market conditions. In fact, it was just an arbitrary price made up to illustrate the point. To eliminate the opportunity to earn the \$2 profit, the forward price should be \$52, which is equal, not coincidentally, to the amount owed on the loan. It is also no coincidence that \$52 is the price of the asset increased by the rate of interest. We will cover this point later in the curriculum, but for now consider that you have just seen your first derivative pricing model.[18]

Of course, many market participants would do this transaction as long as it generated an arbitrage profit. These forces of arbitrage would either push the forward price down or the stock price up, or both, until an equilibrium is reached that eliminates the opportunity to profit at no risk with no commitment of one's own funds.

To summarize, the forces of arbitrage in financial markets assure us that the same asset cannot sell for different prices, nor can two equivalent combinations of assets that produce the same results sell for different prices. Realistically, some arbitrage opportunities can exist on a temporary basis, but they will be quickly exploited, bringing relative prices back in line with each other. Other apparent arbitrage opportunities will be too small to warrant exploiting.

Not to be naive, however, we must acknowledge that there is a large industry of people who call themselves arbitrageurs. So, how can such an industry exist if there are no opportunities for riskless profit? One explanation is that most of the arbitrage transactions are more complex than the simple examples used here. Many involve estimating information, which can result in differing opinions. Arbitrage involving options, for example, usually requires an estimate of a stock's volatility. Different

17 A good study suggestion is to try this example with any future stock price. You should get the same result, a \$2 risk-free profit.

18 This illustration is the quick look at forward pricing alluded to in Section 2.

participants have different opinions about the volatility. It is quite possible that the two counterparties trading with each other believe that each is arbitraging against the other.[19]

But more importantly, the absence of arbitrage opportunities is upheld, ironically, only if participants believe that arbitrage opportunities do exist. If traders believe that no opportunities exist to earn arbitrage profits, then traders will not follow market prices and compare those prices with what they ought to be. Thus, eliminating arbitrage opportunities requires that participants be alert in watching for arbitrage opportunities. In other words, strange as it may sound, disbelief and skepticism concerning the absence of arbitrage opportunities are required for the no-arbitrage rule to be upheld.

Markets in which arbitrage opportunities are either nonexistent or quickly eliminated are relatively efficient markets. Recall that efficient markets are those in which it is not possible to consistently earn returns in excess of those that would be fair compensation for the risk assumed. Although abnormal returns can be earned in a variety of ways, arbitrage profits are definitely examples of abnormal returns. Thus, they are the most egregious violations of the principle of market efficiency.

Throughout the derivatives component of the CFA curriculum, we will use the principle of arbitrage as a dominant theme and assume that arbitrage opportunities cannot exist for any significant length of time nor can any one investor consistently capture them. Thus, prices must conform to models that assume no arbitrage. But we do not want to take the absence of arbitrage opportunities so seriously that we give up and believe that arbitrage opportunities never exist. Otherwise, they will arise and someone else will take them. Consider the rule of arbitrage a law that will be broken from time to time but one that holds far more often than not and one that should be understood and respected.

EXAMPLE 9

Arbitrage

1 Which of the following is a result of arbitrage?
   - **A** The law of one price
   - **B** The law of similar prices
   - **C** The law of limited profitability

2 When an arbitrage opportunity exists, what happens in the market?
   - **A** The combined actions of all arbitrageurs force the prices to converge.
   - **B** The combined actions of arbitrageurs result in a locked-limit situation.
   - **C** The combined actions of all arbitrageurs result in sustained profits to all.

3 Which of the following accurately defines arbitrage?
   - **A** An opportunity to make a profit at no risk
   - **B** An opportunity to make a profit at no risk and with the investment of no capital

19 In reality, many of the transactions that arbitrageurs do are not really arbitrage. They are quite speculative. For example, many people call themselves arbitrageurs because they buy companies that are potential takeover targets and sell the companies they think will be the buyers. This transaction is not arbitrage by any stretch of the definition. Some transactions are called "risk arbitrage," but this term is an oxymoron. As an investment professional, you should simply be prepared for such misuses of words, which simply reflect the flexibility of language.

C An opportunity to earn a return in excess of the return appropriate for the risk assumed

4 Which of the following ways best describes how arbitrage contributes to market efficiency?

A Arbitrage penalizes those who trade too rapidly.

B Arbitrage equalizes the risks taken by all market participants.

C Arbitrage improves the rate at which prices converge to their relative fair values.

Solution to 1:

A is correct. Arbitrage forces equivalent assets to have a single price. There is nothing called the law of similar prices or the law of limited profitability.

Solution to 2:

A is correct. Prices converge because of the heavy demand for the cheaper asset and the heavy supply of the more expensive asset. Profits are not sustained, and, in fact, they are eradicated as prices converge. Locked-limit is a condition in the futures market and has nothing to do with arbitrage.

Solution to 3:

B is correct. An opportunity to profit at no risk could merely describe the purchase of a risk-free asset. An opportunity to earn a return in excess of the return appropriate for the risk assumed is a concept studied in portfolio management and is often referred to as an abnormal return. It is certainly desirable but is hardly an arbitrage because it requires the assumption of risk and the investment of capital. Arbitrage is risk free and requires no capital because selling the overpriced asset produces the funds to buy the underpriced asset.

Solution to 4:

C is correct. Arbitrage imposes no penalties on rapid trading; in fact, it tends to reward those who trade rapidly to take advantage of arbitrage opportunities. Arbitrage has no effect of equalizing risk among market participants. Arbitrage does result in an acceleration of price convergence to fair values relative to instruments with equivalent payoffs.

SUMMARY

This first reading on derivatives introduces you to the basic characteristics of derivatives, including the following points:

- A derivative is a financial instrument that derives its performance from the performance of an underlying asset.
- The underlying asset, called the underlying, trades in the cash or spot markets and its price is called the cash or spot price.
- Derivatives consist of two general classes: forward commitments and contingent claims.
- Derivatives can be created as standardized instruments on derivatives exchanges or as customized instruments in the over-the-counter market.

- Exchange-traded derivatives are standardized, highly regulated, and transparent transactions that are guaranteed against default through the clearinghouse of the derivatives exchange.
- Over-the-counter derivatives are customized, flexible, and more private and less regulated than exchange-traded derivatives, but are subject to a greater risk of default.
- A forward contract is an over-the-counter derivative contract in which two parties agree that one party, the buyer, will purchase an underlying asset from the other party, the seller, at a later date and at a fixed price they agree upon when the contract is signed.
- A futures contract is similar to a forward contract but is a standardized derivative contract created and traded on a futures exchange. In the contract, two parties agree that one party, the buyer, will purchase an underlying asset from the other party, the seller, at a later date and at a price agreed on by the two parties when the contract is initiated. In addition, there is a daily settling of gains and losses and a credit guarantee by the futures exchange through its clearinghouse.
- A swap is an over-the-counter derivative contract in which two parties agree to exchange a series of cash flows whereby one party pays a variable series that will be determined by an underlying asset or rate and the other party pays either a variable series determined by a different underlying asset or rate or a fixed series.
- An option is a derivative contract in which one party, the buyer, pays a sum of money to the other party, the seller or writer, and receives the right to either buy or sell an underlying asset at a fixed price either on a specific expiration date or at any time prior to the expiration date.
- A call is an option that provides the right to buy the underlying.
- A put is an option that provides the right to sell the underlying.
- Credit derivatives are a class of derivative contracts between two parties, the credit protection buyer and the credit protection seller, in which the latter provides protection to the former against a specific credit loss.
- A credit default swap is the most widely used credit derivative. It is a derivative contract between two parties, a credit protection buyer and a credit protection seller, in which the buyer makes a series of payments to the seller and receives a promise of compensation for credit losses resulting from the default of a third party.
- An asset-backed security is a derivative contract in which a portfolio of debt instruments is assembled and claims are issued on the portfolio in the form of tranches, which have different priorities of claims on the payments made by the debt securities such that prepayments or credit losses are allocated to the most-junior tranches first and the most-senior tranches last.
- Derivatives can be combined with other derivatives or underlying assets to form hybrids.
- Derivatives are issued on equities, fixed-income securities, interest rates, currencies, commodities, credit, and a variety of such diverse underlyings as weather, electricity, and disaster claims.
- Derivatives facilitate the transfer of risk, enable the creation of strategies and payoffs not otherwise possible with spot assets, provide information about the spot market, offer lower transaction costs, reduce the amount of capital required, are easier than the underlyings to go short, and improve the efficiency of spot markets.

- Derivatives are sometimes criticized for being a form of legalized gambling and for leading to destabilizing speculation, although these points can generally be refuted.
- Derivatives are typically priced by forming a hedge involving the underlying asset and a derivative such that the combination must pay the risk-free rate and do so for only one derivative price.
- Derivatives pricing relies heavily on the principle of storage, meaning the ability to hold or store the underlying asset. Storage can incur costs but can also generate cash, such as dividends and interest.
- Arbitrage is the condition that two equivalent assets or derivatives or combinations of assets and derivatives sell for different prices, leading to an opportunity to buy at the low price and sell at the high price, thereby earning a risk-free profit without committing any capital.
- The combined actions of arbitrageurs bring about a convergence of prices. Hence, arbitrage leads to the law of one price: Transactions that produce equivalent results must sell for equivalent prices.

PRACTICE PROBLEMS

1 A derivative is *best* described as a financial instrument that derives its performance by:
 A passing through the returns of the underlying.
 B replicating the performance of the underlying.
 C transforming the performance of the underlying.

2 Derivatives are similar to insurance in that both:
 A have an indefinite life span.
 B allow for the transfer of risk from one party to another.
 C allow for the transformation of the underlying risk itself.

3 A beneficial opportunity created by the derivatives market is the ability to:
 A adjust risk exposures to desired levels.
 B generate returns proportional to movements in the underlying.
 C simultaneously take long positions in multiple highly liquid fixed-income securities.

4 Compared with exchange-traded derivatives, over-the-counter derivatives would *most likely* be described as:
 A standardized.
 B less transparent.
 C more transparent.

5 Exchange-traded derivatives are:
 A largely unregulated.
 B traded through an informal network.
 C guaranteed by a clearinghouse against default.

6 The clearing and settlement process of an exchange-traded derivatives market:
 A provides a credit guarantee.
 B provides transparency and flexibility.
 C takes longer than that of most securities exchanges.

7 Which of the following statements *best* portrays the full implementation of post-financial-crisis regulations in the OTC derivatives market?
 A Transactions are no longer private.
 B Most transactions need to be reported to regulators.
 C All transactions must be cleared through central clearing agencies.

8 A characteristic of forward commitments is that they:
 A provide linear payoffs.
 B do not depend on the outcome or payoff of an underlying asset.
 C provide one party the right to engage in future transactions on terms agreed on in advance.

9 In contrast to contingent claims, forward contracts:
 A have their prices chosen by the participants.
 B could end in default by either party.

C can be exercised by physical or cash delivery.

10 Which of the following statements *best* describes the payoff from a forward contract?

A The buyer has more to gain going long than the seller has to lose going short.

B The buyer profits if the price of the underlying at expiration exceeds the forward price.

C The gains from owning the underlying versus owning the forward contract are equivalent.

11 Which of the following statements regarding the settlement of forward contracts is correct?

A Contract settlement by cash has different economic effects from those of a settlement by delivery.

B Non-deliverable forwards and contracts for differences have distinct settlement procedures.

C At cash settlement, when the long party acquires the asset in the market, it effectively pays the forward price.

12 A futures contract is *best* described as a contract that is:

A standardized.

B subject to credit risk.

C marked to market throughout the trading day.

13 Which of the following statements explains a characteristic of futures price limits? Price limits:

A help the clearinghouse manage its credit exposure.

B can typically be expanded intra-day by willing traders.

C establish a band around the final trade of the previous day.

14 Which of the following statements describes an aspect of margin accounts for futures?

A The maintenance margin is always less than the initial margin.

B The initial margin required is typically at least 10% of the futures price.

C A margin call requires a deposit sufficient to raise the account balance to the maintenance margin.

15 Which of the following factors is shared by forwards and futures contracts?

A Timing of profits

B Flexible settlement arrangements

C Nearly equivalent profits by expiration

16 Which of the following derivatives is classified as a contingent claim?

A Futures contracts

B Interest rate swaps

C Credit default swaps

17 In contrast to contingent claims, forward commitments provide the:

A right to buy or sell the underlying asset in the future.

B obligation to buy or sell the underlying asset in the future.

C promise to provide credit protection in the event of default.

18 Which of the following derivatives provide payoffs that are non-linearly related to the payoffs of the underlying?

A Options
B Forwards
C Interest-rate swaps

19 An interest rate swap is a derivative contract in which:
A two parties agree to exchange a series of cash flows.
B the credit seller provides protection to the credit buyer.
C the buyer has the right to purchase the underlying from the seller.

20 Forward commitments subject to default are:
A forwards and futures.
B futures and interest rate swaps.
C interest rate swaps and forwards.

21 A swap is:
A more like a forward than a futures contract.
B subject to simultaneous default by both parties.
C based on an exchange of two series of fixed cash flows.

22 A plain vanilla interest rate swap is also known as:
A a basis swap.
B a fixed-for-floating swap.
C an overnight indexed swap.

23 The notional principal of a swap is:
A not exchanged in the case of an interest rate swap.
B a fixed amount whenever it is matched with a loan.
C equal to the amount owed by one swap party to the other.

24 Which of the following derivatives is *least likely* to have a value of zero at initiation of the contract?
A Futures
B Options
C Forwards

25 The buyer of an option has a contingent claim in the sense that the option creates:
A a right.
B an obligation.
C a linear payoff with respect to gains and losses of the underlying.

26 Which of the following options grants the holder the right to purchase the underlying prior to expiration?
A American-style put option
B European-style call option
C American-style call option

27 A credit derivative is a derivative contract in which the:
A clearinghouse provides a credit guarantee to both the buyer and the seller.
B seller provides protection to the buyer against the credit risk of a third party.
C the buyer and seller provide a performance bond at initiation of the contract.

28 The junior and senior tranches of an asset-backed security:

A have equivalent expected returns.

B have claims on separate underlying portfolios.

C may be differentially impacted by prepayments or credit losses.

29 In a declining interest rate environment, compared with a CMO's Class A tranche, its Class C tranche will be repaid:

A earlier.

B at the same pace.

C later.

30 For a given CDO, which of the following tranches is *most likely* to have the highest expected return?

A Equity

B Senior

C Mezzanine

31 Which of the following derivatives allows an investor to pay the return on a stock index and receive a fixed rate?

A Equity swap

B Stock warrant

C Index futures contract

32 Which of the following is *most likely* the underlying of a plain vanilla interest rate swap?

A 180-day Libor

B 10-year US Treasury bond

C Bloomberg Barclay's US Aggregate Bond Index

33 Currency swaps are:

A rarely used.

B commonly used to manage interest rate risk.

C executed by two parties making a series of interest rate payments in the same currency.

34 Which of the following statements regarding commodity derivatives is correct?

A The primary commodity derivatives are futures.

B Commodities are subject to a set of well-defined risk factors.

C Commodity traders and financial traders today are distinct groups within the financial world.

35 Compared with the underlying spot market, derivative markets are *more likely* to have:

A greater liquidity.

B higher transaction costs.

C higher capital requirements.

36 Which of the following characteristics is *least likely* to be a benefit associated with using derivatives?

A More effective management of risk

B Payoffs similar to those associated with the underlying

C Greater opportunities to go short compared with the spot market

37 Which of the following statements *best* represents information discovery in the futures market?

- **A** The futures price is predictive.
- **B** Information flows more slowly into the futures market than into the spot market.
- **C** The futures market reveals the price that the holder of the asset can take to avoid uncertainty.

38 The derivative markets tend to:

- **A** transfer liquidity from the broader financial markets.
- **B** not reflect fundamental value after it is restored in the underlying market.
- **C** offer a less costly way to exploit mispricing in comparison to other free and competitive financial markets.

39 Which of the following statements *most likely* contributes to the view that derivatives have some role in causing financial crashes?

- **A** Derivatives are the primary means by which leverage and related excessive risk is brought into financial markets.
- **B** Growth in the number of investors willing to speculate in derivatives markets leads to excessive speculative trading.
- **C** Restrictions on derivatives, such as enhanced collateral requirements and credit mitigation measures, in the years leading up to crashes introduce market rigidity.

40 In contrast to gambling, derivatives speculation:

- **A** has a positive public image.
- **B** is a form of financial risk taking.
- **C** benefits the financial markets and thus society.

41 Derivatives may contribute to financial contagion because of the:

- **A** centrally cleared nature of OTC derivatives.
- **B** associated significant costs and high capital requirements.
- **C** reliance by derivatives speculators on large amounts of leverage.

42 The complex nature of derivatives has led to:

- **A** reliable financial models of derivatives markets.
- **B** widespread trust in applying scientific principles to derivatives.
- **C** financial industry employment of mathematicians and physicists.

43 Which of the following is *most likely* to be a destabilizing consequence of speculation using derivatives?

- **A** Increased defaults by speculators and creditors
- **B** Market price swings resulting from arbitrage activities
- **C** The creation of trading strategies that result in asymmetric performance

44 The law of one price is *best* described as:

- **A** the true fundamental value of an asset.
- **B** earning a risk-free profit without committing any capital.
- **C** two assets that will produce the same cash flows in the future must sell for equivalent prices.

45 Arbitrage opportunities exist when:

- **A** two identical assets or derivatives sell for different prices.
- **B** combinations of the underlying asset and a derivative earn the risk-free rate.
- **C** arbitrageurs simultaneously buy takeover targets and sell takeover acquirers.

For questions 46–49, consider a call option selling for $4 in which the exercise price is $50

46 Determine the value at expiration and the profit for a *buyer* if the price of the underlying at expiration is $55.

A $5

B $1

C –$1

47 Determine the value at expiration and the profit for a *buyer* if the price of the underlying at expiration is $48.

A –$4

B $0

C $2

48 Determine the value at expiration and the profit for a *seller* if the price of the underling at expiration is $49.

A $4

B $0

C –$1

49 Determine the value at expiration and the profit for a *seller* if the price of the underling at expiration is $52.

A –$2

B $5

C $2

For questions 50–52, consider the following scenario

Suppose you believe that the price of a particular underlying, currently selling at $99, is going to increase substantially in the next six months. You decide to purchase a call option expiring in six months on this underlying. The call option has an exercise price of $105 and sells for $7.

50 Determine the profit if the price of the underlying six months from now is $99.

A $6

B $0

C –$7

51 Determine the profit if the price of the underlying six months from now is $112.

A $7

B $0

C –$3

52 Determine the profit if the price of the underlying six months from now is $115.

A $0

B $3

C –$3

For questions 53–55, consider the following scenario

Suppose you believe that the price of a particular underlying, currently selling at $99, is going to decrease substantially in the next six months. You decide to purchase a put option expiring in six months on this underlying. The put option has an exercise price of $95 and sells for $5.

53 Determine the profit for you if the price of the underlying six months from now is $100.

A $0

B $5

C –$5

54 Determine the profit for you if the price of the underlying six months from now is $95.

A $0

B $5

C –$5

55 Determine the profit for you if the price of the underlying six months from now is $85.

A $10

B $5

C $0

SOLUTIONS

1 C is correct. A derivative is a financial instrument that transforms the performance of the underlying. The transformation of performance function of derivatives is what distinguishes it from mutual funds and exchange traded funds that pass through the returns of the underlying.

A is incorrect because derivatives, in contrast to mutual funds and exchange traded funds, do not simply pass through the returns of the underlying at payout. B is incorrect because a derivative transforms rather than replicates the performance of the underlying.

2 B is correct. Insurance is a financial contract that provides protection against loss. The party bearing the risk purchases an insurance policy, which transfers the risk to the other party, the insurer, for a specified period of time. The risk itself does not change, but the party bearing it does. Derivatives allow for this same type of risk transfer.

A is incorrect because derivatives, like insurance, have a definite, as opposed to indefinite, life span and expire on a specified date.

C is incorrect because both derivatives and insurance allow for the transfer of risk from one party (the purchaser of the insurance policy or of a derivative) to another party (the insurer or a derivative seller), for a specified period of time. The risk itself does not change, but the party bearing it does.

3 A is correct. Derivatives allow market participants to practice more effective risk management, a process by which an organization, or individual, defines the level of risk it wishes to take, measures the level of risk it is taking, and adjusts the latter to equal the former.

B is incorrect because derivatives are characterized by a relatively high degree of leverage, meaning that participants in derivatives transactions usually have to invest only a small amount, as opposed to a large amount, of their own capital relative to the value of the underlying. This allows participants to generate returns that are disproportional, as opposed to proportional, to movements in the underlying.

C is incorrect because derivatives are not needed to copy strategies that can be implemented with the underlying on a standalone basis. Rather, derivatives can be used to create strategies that cannot be implemented with the underlying alone. Simultaneously taking long positions in multiple highly liquid fixed-income securities is a strategy that can be implemented with the underlying securities on a standalone basis.

4 B is correct. Over-the counter-derivatives markets are customized and mostly unregulated. As a result, over-the-counter markets are less transparent in comparison with the high degree of transparency and standardization associated with exchange-traded derivative markets.

A is incorrect because exchange-traded derivatives are standardized, whereas over-the counter derivatives are customized. C is incorrect because exchange-traded derivatives are characterized by a high degree of transparency because all transactions are disclosed to exchanges and regulatory agencies, whereas over-the-counter derivatives are relatively opaque.

5 C is correct. Exchanged-traded derivatives are guaranteed by a clearinghouse against default.

A is incorrect because traded derivatives are characterized by a relatively high degree of regulation. B is incorrect because the terms of exchange-traded derivatives terms are specified by the exchange.

6 A is correct. The clearing and settlement process of derivative transactions provides a credit guarantee.

B is incorrect because although the exchange markets are said to have transparency, they also involve standardization. That entails a loss of flexibility, with participants limited to only those transactions permitted on the exchange.

C is incorrect because derivatives exchanges clear and settle all contracts overnight, which is faster than most securities exchanges, which require two business days.

7 B is correct. With full implementation of these regulations in the OTC derivatives market, most OTC transactions need to be reported to regulators.

A is incorrect because although under full implementation of the regulations information on most OTC transactions needs to be reported to regulators, many transactions retain a degree of privacy with lower transparency.

C is incorrect because although under full implementation of new regulations a number of OTC transactions have to be cleared through central clearing agencies, there are exemptions that cover a significant percentage of derivative transactions.

8 A is correct because forward commitments provide linear payoffs.

B is incorrect because forward commitments depend on the outcome or payoff of an underlying asset.

C is incorrect because forward commitments obligate parties to make (not provide the right to engage) a final payment contingent on the performance of the underlying.

9 B is correct. In a forward contract, either party could default, whereas in a contingent claim, default is possible only from the short to the long.

A is incorrect because the forward price is set in the pricing of the contract such that the starting contract value is zero, unlike contingent claims, under which parties can select any starting value.

C is incorrect because both forward contracts and contingent claims can be settled by either physical or cash delivery.

10 B is correct. The buyer is obligated to pay the forward price $F_0(T)$ at expiration and receives an asset worth S_T, the price of the underlying. The contract effectively pays off $S_T - F_0(T)$, the value of the contract at expiration. The buyer therefore profits if $S_T > F_0(T)$.

A is incorrect because the long and the short are engaged in a zero-sum game. This is a type of competition in which one participant's gains are the other's losses, with their payoffs effectively being mirror images.

C is incorrect because although the gain from owning the underlying and the gain from owning the forward are both driven by S_T, the price of the underlying at expiration, they are not the same value. The gain from owning the underlying would be $S_T - S_0$, the change in its price, whereas the gain from owning the forward would be $S_T - F_0(T)$, the value of the forward at expiration.

11 C is correct. In the case of cash settlement, the long can acquire the asset, effectively paying the forward price, $F_0(T)$.

A is incorrect because forward contracts settled by cash or by delivery have the same economic effect.

B is incorrect because both non-deliverable forwards and contracts for differences can settle by an exchange of cash.

12 A is correct. A futures contract is a standardized derivative contract.

B is incorrect because through its clearinghouse the futures exchange provides a credit guarantee that it will make up a loss in the event a losing party cannot pay.

C is incorrect because a futures contract is marked to market at the end of each day, a process in which the futures clearinghouse determines an average of the final futures trade of the day and designates that price as the settlement price.

13 A is correct. Price limits are important in helping the clearinghouse manage its credit exposure. Sharply moving prices make it more difficult for the clearinghouse to collect from parties losing money.

B is incorrect because typically the exchange rules allow for an expansion of price limits the next day (not intra-day) if traders are willing.

C is incorrect because price limits establish a band relative to the previous day's settlement price (not final trade).

14 A is correct. The maintenance margin is always significantly lower than the initial margin.

B is incorrect because the initial margin required is typically at most (not at least) 10% of the futures price.

C is incorrect because a margin call requires a deposit large enough to bring the balance up to the initial (not maintenance) margin.

15 C is correct. Comparing the derivatives, forward and futures contracts have nearly equivalent profits by the time of expiration of the forward.

A is incorrect because the timing of profits for a futures contract is different from that of forwards. Forwards realize the full amount at expiration, whereas futures contracts realize their profit in parts on a day-to-day basis.

B is incorrect because the settlement arrangements for the forwards can be agreed on at initiation and written in the contract based on the desires of the engaged parties. However, in the case of a futures contract, the exchange (not the engaged parties) specifies whether physical delivery or cash settlement applies.

16 C is correct. A credit default swap (CDS) is a derivative in which the credit protection seller provides protection to the credit protection buyer against the credit risk of a separate party. CDS are classified as a contingent claim.

A is incorrect because futures contracts are classified as forward commitments. B is incorrect because interest rate swaps are classified as forward commitments.

17 B is correct. Forward commitments represent an obligation to buy or sell the underlying asset at an agreed upon price at a future date.

A is incorrect because the right to buy or sell the underlying asset is a characteristic of contingent claims, not forward commitments. C is incorrect because a credit default swap provides a promise to provide credit protection to the credit protection buyer in the event of a credit event such as a default or credit downgrade and is classified as a contingent claim.

18 A is correct. Options are classified as a contingent claim which provides payoffs that are non-linearly related to the performance of the underlying.

B is incorrect because forwards are classified as a forward commitment, which provides payoffs that are linearly related to the performance of the underlying. C is incorrect because interest-rate swaps are classified as a forward commitment, which provides payoffs that are linearly related to the performance of the underlying.

19 A is correct. An interest rate swap is defined as a derivative in which two parties agree to exchange a series of cash flows: One set of cash flows is variable, and the other set can be variable or fixed.

B is incorrect because a credit derivative is a derivative contract in which the credit protection seller provides protection to the credit protection buyer. C is incorrect because a call option gives the buyer the right to purchase the underlying from the seller.

20 C is correct. Interest rate swaps and forwards are over-the-counter contracts that are privately negotiated and are both subject to default. Futures contracts are traded on an exchange, which provides a credit guarantee and protection against default.

A is incorrect because futures are exchange-traded contracts which provide daily settlement of gains and losses and a credit guarantee by the exchange through its clearinghouse. B is incorrect because futures are exchange-traded contracts which provide daily settlement of gains and losses and a credit guarantee by the exchange through its clearinghouse.

21 A is correct. A swap is a bit more like a forward contract than a futures contract in that it is an OTC contract, so it is privately negotiated and subject to default.

B is incorrect because in a swap, although either party can default, only one party can do so at a particular time. Money owed is based on the net owed by one party to the other, and only the party owing the greater amount can default to the counterparty owing the lesser amount.

C is incorrect because a swap involves an exchange between parties in which at least one party pays a variable series of cash flows determined by an underlying asset or rate.

22 B is correct. A plain vanilla swap is a fixed-for-floating interest rate swap, which is the most common type of swap.

A is incorrect because a basis swap is a transaction based on the TED spread (T-bills versus Eurodollars) and is not the same as a plain vanilla swap.

C is incorrect because an overnight indexed swap is a swap that is tied to a federal funds type of rate, reflecting the rate at which banks borrow overnight, and is not the same as a plain vanilla swap.

23 A is correct. The notional principal of a swap is not exchanged in the case of an interest rate swap.

B is incorrect because an amortizing loan will be matched with a swap with a pre-specified declining (not fixed) notional principal that matches the loan balance.

C is incorrect because the notional principal is equal to the loan balance. Although the loan has an actual balance (the amount owed by the borrower to the creditor), the swap does not have such a balance owed by one swap party to the other.

24 B is correct. The buyer of the option pays the option premium to the seller of the option at the initiation of the contract. The option premium represents the value of the option, whereas futures and forwards have a value of zero at the initiation of the contract.

A is incorrect because no money changes hands between parties at the initiation of the futures contract, thus the value of the futures contract is zero at initiation. C is incorrect because no money changes hands between parties at the initiation of the forward contract, thus the value of the forward contract is zero at initiation.

25 A is correct. A contingent claim, a derivative in which the outcome or payoff depends on the outcome or payoff of an underlying asset, has come to be associated with a right, but not an obligation, to make a final payment contingent on the performance of the underlying.

B is incorrect because an option, as a contingent claim, grants the right but not the obligation to buy or sell the underlying at a later date.

C is incorrect because the holder of an option has a choice of whether to exercise the option. This choice creates a payoff that transforms the underlying payoff in a more pronounced manner than does a forward, futures, or swap, which provide linear payoffs. Options are different in that they limit losses in one direction.

26 C is correct. The right to buy the underlying is referred to as a call option. Furthermore, options that can be exercised prior to the expiration date are referred to as American-style options.

A is incorrect because a put option grants the holder the right to sell, as opposed to buy, the underlying.

B is incorrect because European-style options can only be exercised at expiration.

27 B is correct. A credit derivative is a derivative contract in which the credit protection seller provides protection to the credit protection buyer against the credit risk of a third party.

A is incorrect because the clearinghouse provides a credit guarantee to both the buyer and the seller of a futures contract, whereas a credit derivative is between two parties, in which the credit protection seller provides a credit guarantee to the credit protection buyer. C is incorrect because futures contracts require that both the buyer and the seller of the futures contract provide a cash deposit for a portion of the futures transaction into a margin account, often referred to as a performance bond or good faith deposit.

28 C is correct. An asset-backed security is a derivative contract in which a portfolio of debt instruments is assembled and claims are issued on the portfolio in the form of tranches, which have different priorities of claims on the payments made by the debt securities such that prepayments or credit losses are allocated to the most junior tranches first and the most senior tranches last.

A is incorrect because the expected returns of the tranches vary according to the perceived credit risk, with the senior tranches having the highest credit quality and the junior tranches the lowest. Thus, the senior tranches have the lowest expected returns and the junior tranches have the highest. Notably, in a bond mutual fund or an ETF, all investors in the fund have equal claims, and so the rate of return earned by each investor is the same.

B is incorrect because an asset-backed security is a derivative contract in which a single portfolio of securities is assembled and claims are issued on the portfolio in the form of tranches.

29 A is correct. Lower interest rates entice homeowners to pay off their mortgages early because they can refinance at lower rates. The most junior tranche in a CMO will bear the first wave of prepayments until that tranche has been completely repaid its full principal investment. At that point, the next tranche will bear prepayments until that tranche has been fully repaid. Therefore, the Class C tranche of a CMO will be repaid before the more senior Class A tranche.

B is incorrect because the tranches, which have different priorities of claims on the principal payments made by the underlying mortgages, will see prepayments allocated to the most junior tranches first and the most senior tranches last.

C is incorrect because the most junior tranche in a CMO will bear the first wave of prepayments until that tranche has been completely repaid its full principal investment. At that point, the next tranche will bear prepayments until that tranche has been fully repaid. Therefore, the Class C tranche will be repaid prior to, not after, the Class A tranche.

30 A is correct. The expected returns of the tranches vary according to the perceived credit risk, with the senior tranches having the highest credit quality and the junior tranches the lowest. Thus, the senior tranches have the lowest expected returns and the junior tranches have the highest. The most junior tranche is sometimes called the "equity tranche."

B is incorrect because the senior tranches in a CDO have the lowest expected returns and the junior (or equity) tranches have the highest.

C is incorrect because the senior tranches in a CDO have the lowest expected returns and the junior (or equity) tranches have the highest. A mezzanine tranche is intermediate between the senior and junior tranches.

31 A is correct. Equity swaps, also known as index swaps, are quite popular and permit investors to pay the return on one stock index and receive the return on another index or a fixed rate.

B is incorrect because warrants are options that are sold directly to the public, allowing holders to exercise and buy shares directly from the company as opposed to using stock indexes to determine returns.

C is incorrect because although index derivatives in the form of options, forwards, futures, and swaps are very popular, paying the return on a stock index and receiving a fixed rate describes an equity swap (or index swap), not a futures contract.

32 A is correct. In a plain vanilla interest rate swap, an interest rate, such as Libor, serves as the underlying. A plain vanilla interest rate swap is one of many derivatives in which a rate, not the instrument that pays the rate, is the underlying.

B is incorrect because a plain vanilla interest rate swap is one of many derivatives in which a rate, not an instrument that pays a rate, is the underlying.

C is incorrect because a plain vanilla interest rate swap is one of many derivatives in which a rate, not an instrument (or index) that pays a rate, is the underlying.

33 B is correct. Because interest rates and currencies are both subject to change, a currency swap has two sources of risk. Furthermore, companies operating across borders are subject to both interest rate risk and currency risk, and currency swaps are commonly used to manage these risks.

A is incorrect because currency risk is a major factor in global financial markets, and the currency derivatives market is extremely large, as opposed to small.

C is incorrect because a currency swap is executed by two parties making a series of interest rate payments to each other in different currencies, as opposed to the same currency.

34 A is correct. The primary commodity derivatives are futures, but forwards, swaps, and options are also used.

B is incorrect because the commodity market is extremely large and subject to an almost unimaginable array of risks.

C is incorrect because commodity and financial traders have become relatively homogeneous since the creation of financial futures. Historically, commodity traders and financial traders were quite different groups, and there used to be a tendency to think of the commodity world as somewhat separate from the financial world.

35 A is correct. Derivative markets typically have greater liquidity than the underlying spot market as a result of the lower capital required to trade derivatives compared with the underlying. Derivatives also have lower transaction costs and lower capital requirements than the underlying.

B is incorrect because transaction costs for derivatives are lower than the underlying spot market. C is incorrect because derivatives markets have lower capital requirements than the underlying spot market.

36 B is correct. One of the benefits of derivative markets is that derivatives create trading strategies not otherwise possible in the underlying spot market, thus providing opportunities for more effective risk management than simply replicating the payoff of the underlying.

A is incorrect because effective risk management is one of the primary purposes associated with derivative markets. C is incorrect because one of the operational advantages associated with derivatives is that it is easier to go short compared to the underlying spot market.

37 C is correct. The futures market reveals the price that the holder of an asset could take and avoid the risk of uncertainty.

A is incorrect because although the futures price is sometimes thought of as predictive, it provides only a little more information than does a spot price and is not really a forecast of the futures spot price.

B is incorrect because by virtue of the fact that the futures market requires less capital, information can flow into the futures market before it gets into the spot market.

38 C is correct. When prices deviate from fundamental values, derivative markets offer a less costly way to exploit mispricing in comparison to other free and competitive financial markets.

A is incorrect because derivative markets tend to transfer liquidity to (not from) the broader financial markets, because investors are far more willing to trade if they can more easily manage their risk, trade at lower cost and with less capital, and go short more easily. An increased willingness to trade leads to a more liquid market.

B is incorrect because it is likely (not unlikely) that fundamental value will be reflected in the derivative markets both before and after it is restored in the underlying market owing to lower capital requirements and transaction costs in the derivative markets.

39 B is correct. Opponents of derivatives claim that excessive speculative trading brings instability to the markets. Defaults by speculators can lead to defaults by their creditors, their creditors' creditors, and so on.

A is incorrect because derivatives are one of many mechanisms through which excessive risk can be taken. There are many ways to take on leverage that look far less harmful but can be just as risky.

C is incorrect because responses to crashes and crises typically call for more rules and regulations restricting the use of derivatives, such as requiring more collateral and credit mitigation measures. Such rules and regulations are generally implemented after a crash and are directed at limiting government bailouts of the costs from derivatives risks.

40 C is correct. Derivatives trading brings extensive benefits to financial markets (low costs, low capital requirements, ease of going short, etc.) and thus benefits society as a whole. Gambling, on the other hand, typically benefits only a limited number of participants.

A is incorrect because the general image of speculators is not a good one. Speculators are often thought to be short-term traders who attempt to exploit temporary inefficiencies, caring little about long-term fundamental values.

B is incorrect because speculation and gambling are both forms of financial risk taking.

41 C is correct. Opponents argue that speculators use large amounts of leverage, thereby subjecting themselves and their creditors to substantial risk if markets do not move in their hoped-for direction. Defaults by speculators can then lead to defaults by their creditors, their creditors' creditors, and so on. These effects can, therefore, be systemic and reflect an epidemic contagion whereby instability can spread throughout markets and an economy, if not the entire world.

A is incorrect because central clearing of OTC derivatives, similar to how exchange-traded derivatives are cleared, is intended to lessen the risk of contagion.

B is incorrect because it is derivatives' low cost and low capital requirements, not high cost and high capital requirements, that opponents point to as contributing to an excessive amount of speculative trading that brings instability to the markets.

42 C is correct. Many derivatives are extremely complex and require a high-level understanding of mathematics. As a result, the financial industry employs many mathematicians, physicists, and computer scientists.

A is incorrect because scientists create models of markets by using scientific principles that often fail. For example, to a physicist modeling the movements of celestial bodies, the science is reliable and the physicist is unlikely to misapply the science. The same science applied to financial markets is far less reliable. Financial markets are driven by the actions of people who are not as consistent as the movements of celestial bodies.

B is incorrect because the complex nature of derivatives has made many distrust, as opposed to trust, derivatives, the people who work with them, and the scientific methods they use.

43 A is correct. The benefits of derivatives, such as low transaction costs, low capital requirements, use of leverage, and the ease in which participants can go short, also can result in excessive speculative trading. These activities can lead to defaults on the part of speculators and creditors.

B is incorrect because arbitrage activities tend to bring about a convergence of prices to intrinsic value. C is incorrect because asymmetric performance is not itself destabilizing.

44 C is correct. The law of one price occurs when market participants engage in arbitrage activities so that identical assets sell for the same price in different markets.

A is incorrect because the law of one price refers to identical assets. B is incorrect because it refers to arbitrage not the law of one price.

45 A is correct. Arbitrage opportunities exist when the same asset or two equivalent combinations of assets that produce the same results sell for different prices. When this situation occurs, market participants would buy the asset in the cheaper market and simultaneously sell it in the more expensive market, thus earning a riskless arbitrage profit without committing any capital.

B is incorrect because it is not the definition of an arbitrage opportunity. C is incorrect because it is not the definition of an arbitrage opportunity.

46 B is correct. $C_T = Max(0,S_T - X) = Max(0,55 - 50) = 5$

$\Pi = C_T - C_0 = 5 - 4 = 1$

47 A is correct. $C_T = Max(0,S_T - X) = Max(0,48 - 50) = 0$

$\Pi = C_T - C_0 = 0 - 4 = -4$

48 A is correct. $-C_T = -Max(0,S_T - X) = -Max(0,49 - 50) = 0$

$\Pi = -C_T + C_0 = -0 + 4 = 4$

49 C is correct. $-C_T = -Max(0,S_T - X) = -Max(0,52 - 50) = -2$

$\Pi = -C_T + C_0 = -2 + 4 = 2$

50 C is correct. $C_T = Max(0,S_T - X) = Max(0,99 - 105) = 0$

$\Pi = C_T - C_0 = 0 - 7 = -7$

51 B is correct. $C_T = Max(0,S_T - X) = Max(0,112 - 105) = 7$

$\Pi = C_T - C_0 = 7 - 7 = 0$

Note: $112 is the breakeven price

52 B is correct. $C_T = Max(0,S_T - X) = Max(0,115 - 105) = 10$

$\Pi = C_T - C_0 = 10 - 7 = 3$

53 C is correct. $P_T = Max(0, X - S_T) = Max(0,95 - 100) = 0$

$\Pi = P_\mathrm{T} - P_0 = 0 - 5 = -5$

54 C is correct. $P_T = Max(0, X - S_T) = Max(0,95 - 95) = 0$

$\Pi = P_T - P_0 = 0 - 5 = -5$

55 B is correct. $P_T = Max(0, X - S_T) = Max(0,95 - 85) = 10$

$\Pi = P_T - P_0 = 10 - 5 = 5$

READING

46

Basics of Derivative Pricing and Valuation

by Don M. Chance, PhD, CFA

Don M. Chance, PhD, CFA, is at Louisiana State University (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** explain how the concepts of arbitrage, replication, and risk neutrality are used in pricing derivatives;
☐	**b.** explain the difference between value and price of forward and futures contracts;
☐	**c.** calculate a forward price of an asset with zero, positive, or negative net cost of carry;
☐	**d.** explain how the value and price of a forward contract are determined at expiration, during the life of the contract, and at initiation;
☐	**e.** describe monetary and nonmonetary benefits and costs associated with holding the underlying asset and explain how they affect the value and price of a forward contract;
☐	**f.** define a forward rate agreement and describe its uses;
☐	**g.** explain why forward and futures prices differ;
☐	**h.** explain how swap contracts are similar to but different from a series of forward contracts;
☐	**i.** explain the difference between value and price of swaps;
☐	**j.** explain the exercise value, time value, and moneyness of an option;
☐	**k.** identify the factors that determine the value of an option and explain how each factor affects the value of an option;
☐	**l.** explain put–call parity for European options;
☐	**m.** explain put–call–forward parity for European options;
☐	**n.** explain how the value of an option is determined using a one-period binomial model;
☐	**o.** explain under which circumstances the values of European and American options differ.

1 INTRODUCTION

It is important to understand how prices of derivatives are determined. Whether one is on the buy side or the sell side, a solid understanding of pricing financial products is critical to effective investment decision making. After all, one can hardly determine what to offer or bid for a financial product, or any product for that matter, if one has no idea how its characteristics combine to create value.

Understanding the pricing of financial assets is important. Discounted cash flow methods and models, such as the capital asset pricing model and its variations, are useful for determining the prices of financial assets. The unique characteristics of derivatives, however, pose some complexities not associated with assets, such as equities and fixed-income instruments. Somewhat surprisingly, however, derivatives also have some simplifying characteristics. For example, as we will see in this reading, in well-functioning derivatives markets the need to determine risk premiums is obviated by the ability to construct a risk-free hedge. Correspondingly, the need to determine an investor's risk aversion is irrelevant for derivative pricing, although it is certainly relevant for pricing the underlying.

The purpose of this reading is to establish the foundations of derivative pricing on a basic conceptual level. The following topics are covered:

- How does the pricing of the underlying asset affect the pricing of derivatives?
- How are derivatives priced using the principle of arbitrage?
- How are the prices and values of forward contracts determined?
- How are futures contracts priced differently from forward contracts?
- How are the prices and values of swaps determined?
- How are the prices and values of European options determined?
- How does American option pricing differ from European option pricing?

This reading is organized as follows. Sections 2–4 explore two related topics, the pricing of the underlying assets on which derivatives are created and the principle of arbitrage. Sections 5–7 describe the pricing and valuation of forwards, futures, and swaps. Sections 8–12 introduce the pricing and valuation of options. Section 13 provides a summary.

2 BASIC DERIVATIVE CONCEPTS, PRICING THE UNDERLYING

d explain how the value and price of a forward contract are determined at expiration, during the life of the contract, and at initiation;

In this section, we will briefly review the concepts associated with derivatives, the types of derivatives, and the pricing principles of the underlying assets. We will also look at arbitrage, a critical concept that links derivative pricing to the price of the underlying.

2.1 Basic Derivative Concepts

The definition of a derivative is as follows:

> *A derivative is a financial instrument that derives its performance from the performance of an underlying asset.*

A derivative is created as a contract between two parties, the buyer and the seller. Derivatives trade in markets around the world, which include organized exchanges, where highly standardized and regulated versions exist, and over-the-counter markets, where customized and more lightly regulated versions trade. The basic characteristics of derivatives that influence pricing are not particularly related to where the derivatives trade, but are critically dependent on the types of derivatives.

The two principal types of derivatives are forward commitments and contingent claims. A forward commitment is an obligation to engage in a transaction in the spot market at a future date at terms agreed upon today.[1] By entering into a forward commitment, a party locks in the terms of a transaction that he or she will conduct later. The word "commitment" is critical here. A forward contract is a firm obligation.

There are three types of forward commitments: forward contracts, futures contracts, and swap contracts. These contracts can be referred to more simply as forwards, futures, and swaps.

> *A* **forward contract** *is an over-the-counter derivative contract in which two parties agree that one party, the buyer, will purchase an underlying asset from the other party, the seller, at a later date at a fixed price they agree upon when the contract is signed.*

> *A* **futures contract** *is a standardized derivative contract created and traded on a futures exchange in which two parties agree that one party, the buyer, will purchase an underlying asset from the other party, the seller, at a later date at a price agreed upon by the two parties when the contract is initiated and in which there is a daily settling of gains and losses and a credit guarantee by the futures exchange through its clearinghouse.*

> *A* **swap contract** *is an over-the-counter derivative contract in which two parties agree to exchange a series of cash flows whereby one party pays a variable series that will be determined by an underlying asset or rate and the other party pays either 1) a variable series determined by a different underlying asset or rate or 2) a fixed series.*

As these definitions illustrate, forwards and futures are similar. They both establish the terms of a spot transaction that will occur at a later date. Forwards are customized, less transparent, less regulated, and subject to higher counterparty default risk. Futures are standardized, more transparent, more regulated, and generally immune to counterparty default. A swap is equivalent to a series of forward contracts, a point that will be illustrated later.

A contingent claim is a derivative in which the outcome or payoff is determined by the outcome or payoff of an underlying asset, conditional on some event occurring. Contingent claims include options, credit derivatives, and asset-backed securities. Because credit derivatives and asset-backed securities are highly specialized, this reading will focus only on options.

Recall the definition of an option:

> *An* **option** *is a derivative contract in which one party, the buyer, pays a sum of money to the other party, the seller or writer, and receives the right to either buy or sell an underlying asset at a fixed price either on a specific expiration date or at any time prior to the expiration date.*

1 Remember that the term "spot market" refers to the market in which the underlying trades. A transaction in the spot market involves a buyer paying for an asset and receiving it right away or at least within a few days, given the normal time required to settle a financial transaction.

Options can be either customized over-the-counter contracts or standardized and traded on exchanges.

Because derivatives take their prices from the price of the underlying, it is important to first understand how the underlying is priced. We will approach the underlying from a slightly different angle, one that emphasizes the often-subtle costs of holding the underlying, which turn out to play a major role in derivative pricing.

2.2 Pricing the Underlying

The four main types of underlying on which derivatives are based are equities, fixed-income securities/interest rates, currencies, and commodities. Equities, fixed-income securities (but not interest rates), currencies, and commodities are all assets. An interest rate is not an asset, but it can be structured as the underlying of a derivative.[2]

Consider a generic underlying asset. This asset is something of value that you can own. Some assets are financial assets, such as equities, bonds, and currencies, and some are real assets, such as commodities (e.g., gold, oil, and agricultural products) and certain physical objects (e.g., houses, automobiles, and computers).

The price of a financial asset is often determined using a present value of future cash flows approach. The value of the financial asset is the expected future price plus any interim payments such as dividends or coupon interest discounted at a rate appropriate for the risk assumed. Such a definition presumes a period of time over which an investor anticipates holding an asset, known as the holding period. The investor forecasts the price expected to prevail at the end of the holding period as well as any cash flows that are expected to be earned over the holding period. He then takes that predicted future price and expected cash flows and finds their current value by discounting them to the present. Thereby, the investor arrives at a fundamental value for the asset and will compare that value with its current market price. Based on any differential relative to the cost of trading and his confidence in his valuation model, he will make a decision about whether to trade.

2.2.1 *The Formation of Expectations*

Let us first assume that the underlying does not pay interest or dividends, nor does it have any other cash flows attributable to holding the asset. Exhibit 1 illustrates the basic idea behind the valuation process. Using a probability distribution, the investor forecasts the future over a holding period spanning time 0 to time T. The center of the distribution is the expected price of the asset at time T, which we denote as $E(S_T)$, and represents the investor's prediction of the spot price at T. The investor knows there is risk, so this prediction is imperfect—hence the reason for the probability distribution. Nonetheless, at time 0 the investor makes her best prediction of the spot price at time T, which becomes the foundation for determining what she perceives to be the value of the asset.[3]

2 This is a good example of why it is best not to use the term "underlying *asset*" when speaking of derivatives. Not all derivatives have underlying assets, but all have underlyings, some of which are not assets. Some other examples of non-asset underlyings used in derivatives are weather, insurance claims, and shipping rates. There are also some derivatives in which the underlying is another derivative.

3 The distribution shown here is symmetrical and relatively similar to a normal distribution, but this characterization is for illustrative purposes only. We are making no assumptions about symmetry or normality at this point.

Exhibit 1 The Formation of Expectations for an Asset

2.2.2 *The Required Rate of Return on the Underlying Asset*

To determine the value of the asset, this prediction must be converted into its price or present value. The specific procedure is to discount this expected future price, but that is the easy part. Determining the rate at which to discount the expected future price is the hard part. We use the symbol k to denote this currently unknown discount rate, which is often referred to as the required rate of return and sometimes the expected rate of return or just the expected return. At a minimum, that rate will include the risk-free rate of interest, which we denote as r. This rate represents the opportunity cost, or so-called time value of money, and reflects the price of giving up your money today in return for receiving more money later.

2.2.3 *The Risk Aversion of the Investor*

At this point, we must briefly discuss an important characteristic of investors: their degree of risk aversion. We can generally characterize three potential types of investors by how they feel about risk: risk averse, risk neutral, or risk seeking.

Risk-neutral investors are willing to engage in risky investments for which they expect to earn only the risk-free rate. Thus, they do not expect to earn a premium for bearing risk. For risk-averse investors, however, risk is undesirable, so they do not consider the risk-free rate an adequate return to compensate them for the risk. Thus, risk-averse investors require a risk premium, which is an increase in the expected return that is sufficient to justify the acceptance of risk. All things being equal, an investment with a higher risk premium will have a lower price. It is very important to understand, however, that risk premiums are not automatically earned. They are merely expectations. Actual outcomes can differ. Clearly stocks that decline in value did not earn risk premiums, even though someone obviously bought them with the expectation that they would. Nonetheless, risk premiums must exist in the long run or risk-averse investors would not accept the risk.

The third type of investor is one we must mention but do not treat as realistic. Risk seekers are those who prefer risk over certainty and will pay more to invest when there is risk, implying a negative risk premium. We almost always assume that investors prefer certainty over uncertainty, so we generally treat a risk-seeking investor as just a theoretical possibility and not a practical reality.[4]

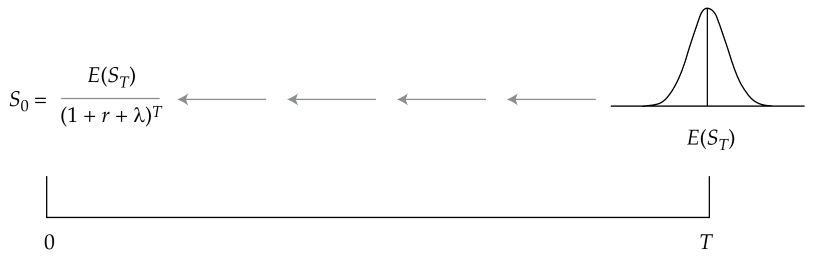
4 People who gamble in casinos or play lotteries appear to be risk-seekers, given the advantage of the casino or the lottery organizer, but they are merely earning utility from the game itself, not necessarily from the expected financial outcome.

We will assume that investors are risk averse. To justify taking risk, risk-averse investors require a risk premium. We will use the Greek symbol λ (lambda) to denote the risk premium.[5]

2.2.4 *The Pricing of Risky Assets*

Exhibit 2 illustrates the process by which an investor obtains the current price, S_0, by discounting the expected future price of an asset with no interim cash flows, $E(S_T)$, by r (the risk-free rate) plus λ (the risk premium) over the period from 0 to T.

Exhibit 2 Discounting the Expected Future Price to Obtain the Current Price



2.2.5 *Other Benefits and Costs of Holding an Asset*

Many assets generate benefits and some incur costs to their owners. Some of these costs are monetary and others are nonmonetary. The dividends paid by companies and coupon interest paid by borrowers on their bonds represent obvious benefits to the holders of these securities. With currencies representing investments that earn the risk-free rate in a foreign country, they too generate benefits in the form of interest. Barring default, interest payments on bonds and currencies are relatively certain, so we will treat them as such. Dividend payments are not certain, but dividends do tend to be fairly predictable. As such, we will make an assumption common to most derivative models that dividends are certain.[6]

There is substantial evidence that some commodities generate a benefit that is somewhat opaque and difficult to measure. This benefit is called the **convenience yield**. It represents a nonmonetary advantage of holding the asset. For most financial assets, convenience yields are either nonexistent or extremely limited. Financial assets do not possess beauty that might make a person enjoy owning them just to look at them. Convenience yields are primarily associated with commodities and generally exist as a result of difficulty in either shorting the commodity or unusually tight supplies. For example, if a commodity cannot be sold short without great difficulty or cost, the holder of the commodity has an advantage if market conditions suggest that the commodity should be sold. Also, if a commodity is in short supply, the holders of the commodity can sometimes extract a price premium that is believed by some to be higher than what would be justified in well-functioning markets. The spot price of the commodity could even be above the market's expectation of its future price, a condition that would seem to imply a negative expected return. This scenario raises the question of why anyone would want to hold the commodity if its expected return is negative. The convenience yield provides a possible explanation that attributes an

---

5 Although the risk-free rate is invariant with a country's economy, the risk premium varies with the amount of risk taken. Thus, while the risk-free rate is the same when applied to every investment, the risk premium is not the same for every investment.

6 Some derivative models incorporate uncertain dividends and interest, but those are beyond the scope of this introductory reading.

implied but non-financial expected return to the advantage of holding a commodity in short supply. The holder of the commodity has the ability to sell it when market conditions suggest that selling is advisable and short selling is difficult.

One cost incurred in owning commodities is the cost of storage. One could hardly own gold, oil, or wheat without incurring some costs in storing these assets. There are also costs incurred in protecting and insuring some commodities against theft or destruction. Depending on the commodity, these costs can be quite significant. For financial assets, however, the storage costs are so low that we can safely ignore them.

Finally, there is the opportunity cost of the money invested. If a person buys an asset, he forgoes interest on his money. The effect on this interest is reflected by compounding the price paid for the asset to a future value at the risk-free rate of interest. Thus, an investor who buys a stock that costs £50 in a market in which the risk-free rate is 4% will effectively have paid £50 × 1.04 = £52 a year later. Of course, the stock could be worth any value at that time, and any gain or loss should be determined in comparison to the effective price paid of £52.

As we described earlier, we determine the current price of an asset by discounting the expected future price by the sum of the risk-free rate (r) plus the risk premium (λ). When we introduce costs and benefits of holding the asset, we have to make an adjustment. With the exception of this opportunity cost of money, we will incorporate the effect of these costs and benefits by determining their value at the end of the holding period. Under the assumption that these costs and benefits are certain, we can then discount them at the risk-free rate to obtain their present value. There is a logic to doing it this way (i.e., finding their future value and discounting back to the present, as opposed to finding their present value directly). By finding their future value, we are effectively saying that the costs and benefits adjust the expected payoff at the end of the holding period. But because they are certain, we can discount their effects at the risk-free rate. So we have effectively just found their present value. The net effect is that the costs reduce the current price and the benefits increase the current price. We use the symbol θ (theta) to denote the present value of the costs and γ (gamma) as the present value of any benefits.

The net of the costs and benefits is often referred to by the term **carry**, or sometimes **cost of carry**. The holding, storing, or "carrying" of an asset is said to incur a net cost that is essentially what it takes to "carry" an asset. Exhibit 3 illustrates the effect in which the carry adjusts the price of an asset in the valuation process.

Exhibit 3 Pricing an Asset That Incurs Costs and Generates Benefits

EXAMPLE 1

Pricing the Spot Asset

1 Which of the following factors does **not** affect the spot price of an asset that has no interim costs or benefits?
   - A The time value of money
   - B The risk aversion of investors
   - C The price recently paid by other investors

2 Which of the following does **not** represent a benefit of holding an asset?
   - A The convenience yield
   - B An optimistic expected outlook for the asset
   - C Dividends if the asset is a stock or interest if the asset is a bond

Solution to 1:

C is correct. The price recently paid by other investors is past information and does not affect the spot price. The time value of money and the risk aversion of investors determine the discount rate. Only current information is relevant as investors look ahead, not back.

Solution to 2:

B is correct. An optimistic forecast for the asset is not a benefit of holding the asset, but it does appear in the valuation of the asset as a high expected price at the horizon date. Convenience yields and dividends and interest are benefits of holding the asset.

To recap, although the various underlyings differ with respect to the specifics of pricing, all of them are based on expectations, risk, and the costs and benefits of holding a specific underlying. Understanding how assets are priced in the spot market is critical to understanding how derivatives are priced. To understand derivative pricing, it is necessary to establish a linkage between the derivative market and the spot market. That linkage occurs through arbitrage.

3 THE PRINCIPLE OF ARBITRAGE

a explain how the concepts of arbitrage, replication, and risk neutrality are used in pricing derivatives;

Arbitrage is a type of transaction undertaken when two assets or portfolios produce identical results but sell for different prices. If a trader buys the asset or portfolio at the cheaper price and sells it at the more expensive price, she will generate a net inflow of funds at the start. Because the two assets or portfolios produce identical results, a long position in one and a short position in the other means that at the end of the holding period, the payoffs offset. Hence, no money is gained or lost at the end of the holding period, so there is no risk. The net effect is that the arbitrageur receives money at the start and never has to pay out any money later. Such a situation amounts to free money, like walking down the street, finding money on the ground, and never having to give it up. Exhibit 4 illustrates this process for assets A and B, which have no dividends or other benefits or costs and pay off identically but sell for different prices, with $S_0^A < S_0^B$.

Exhibit 4 Executing an Arbitrage

Given: Assets A and B produce the same values at time T but at time 0, A is selling for less than B.

$S_0^A < S_0^B$:
Buy A at S_0^A
Sell B at S_0^B
Cash flow = $S_0^B - S_0^A (> 0)$

$S_T^A = S_T^B$:
Sell A for S_T^A
Buy B for S_T^B
Cash flow = $S_T^A - S_T^B (= 0)$

0 T

3.1 The (In)Frequency of Arbitrage Opportunities

When arbitrage opportunities exist, traders exploit them very quickly. The combined actions of many traders engaging in the same transaction of buying the low-priced asset or portfolio and selling the high-priced asset or portfolio results in increased demand and an increasing price for the former and decreased demand and a decreasing price for the latter. This market activity will continue until the prices converge. Assets that produce identical results can thus have only one true market price. This rule is called the "law of one price." With virtually all market participants alert for the possibility of earning such profits at no risk, it should not be surprising that arbitrage opportunities are rare.

In practice, prices need not converge precisely, or even all that quickly, because the transaction cost of exploiting an opportunity could exceed the benefit. For example, say you are walking down the sidewalk of the Champs-Élysées in Paris and notice a €1 coin on the sidewalk. You have a bad back, and it would take some effort to bend over. The transaction cost of exploiting this opportunity without any risk could exceed the benefit of the money. Some arbitrage opportunities represent such small discrepancies that they are not worth exploiting because of transaction costs.

Significant arbitrage opportunities, however, will be exploited. A significant opportunity arises from a price differential large enough to overcome the transaction costs. Any such price differential will continue to be exploited until the opportunity disappears. Thus, if you find a €10 note on the Champs-Élysées sidewalk, there is a good chance you will find it worth picking up (even with your bad back), and even if you do not pick it up, it will probably not be there for long. With enough people alert for such opportunities, only a few will arise, and the ones that do will be quickly exploited and disappear. In this manner, arbitrage makes markets work much more efficiently.

3.2 Arbitrage and Derivatives

It may be difficult to conceive of many investments that would produce identical payoffs. Even similar companies such as McDonalds and Burger King, which are in the same line of business, do not perform identically. Their performance may be correlated, but each has its own unique characteristics. For equity securities and with no derivatives involved, about the only such situation that could exist in reality is a stock that trades simultaneously in two different markets, such as Royal Dutch Shell, which trades in Amsterdam and London but is a single company. Clearly there can be only one price. If those two markets operate in different currencies, the currency-adjusted prices should be the same. Bonds issued by the same borrower are also potentially arbitrageable. All bonds of an issuer will be priced off of the term structure of interest rates. Because of this common factor, bonds of different maturities can be arbitraged against each other. But in general, two securities are unlikely to perform identically.

The picture changes, however, if we introduce derivatives. For most derivatives, the payoffs come (derive) directly from the value of the underlying at the expiration of the derivative. Although no one can predict with certainty the value of the underlying at expiration, as soon as that value is determined, the value of the derivative at expiration becomes certain. So, while the performance of McDonalds' stock may have a strong correlation to the performance of Burger King's stock, neither completely determines the other. But derivatives on McDonalds' stock and derivatives on Burger King's stock are completely determined by their respective stocks. All of the uncertainty in a derivative comes from the uncertainty in the underlying. As a result, the price of the derivative is tied to the price of the underlying. That being the case, the derivative can be used to hedge the underlying, or vice versa.

Exhibit 5 illustrates this point. When a long position in the underlying is combined with a short position in the derivative to produce a perfect hedge, all of the risk is eliminated and the position should earn the risk-free rate. If not, arbitrageurs begin to trade. If the position generates a return in excess of the risk-free rate, the arbitrageurs see an opportunity because the hedged position of the underlying (long asset and short derivative) earns more than the risk-free rate and a risk-free loan undertaken as a borrower incurs a cost equal to the risk-free rate. Therefore, implementing the hedged position and borrowing at the risk-free rate earns a return in excess of the risk-free rate, incurs a cost of the risk-free rate, and has no risk. As a result, an investor can earn excess return at no risk without committing any capital. Arbitrageurs will execute this transaction in large volumes, continuing to exploit the pricing discrepancy until market forces push prices back in line such that both risk-free transactions earn the risk-free rate.

Exhibit 5 Hedging the Underlying with a Derivative (or Vice Versa)

Out of this process, one and only one price can exist for the derivative. Otherwise, there will be an arbitrage opportunity. We typically take the underlying price as given and infer the unique derivative price that prohibits any arbitrage opportunities. Most derivatives pricing models are established on this foundation. We simply assume that no arbitrage opportunities can exist and infer the derivative price that guarantees there are no arbitrage opportunities.

3.3 Arbitrage and Replication

Because a long asset and a short derivative on the asset can be combined to produce a position equivalent to a risk-free bond, it follows that the long asset and a short risk-free asset (meaning to borrow at the risk-free rate) can be combined to produce a long derivative. Alternatively, a short derivative and the short risk-free asset can be combined to produce a short asset position. Exhibit 6 shows this process, referred to as **replication**. Replication is the creation of an asset or portfolio from another asset, portfolio, and/or derivative.

Exhibit 6 Arbitrage, Replication, and Derivatives

Long asset	+	Short derivative	=	Long risk-free asset (lending)
Long asset	+	Short risk-free asset (borrowing)	=	Long derivative
Short derivative	+	Short risk-free asset (borrowing)	=	Short asset

If all assets are correctly priced to prohibit arbitrage, however, the ability to replicate seems useless. Why would one replicate an asset or derivative if there is no cost advantage? Buying a government security to earn the risk-free rate is easier than buying the asset and selling a derivative to produce a risk-free position. At this point, that is certainly a reasonable question. As we progress through this material, however, we will relax the assumption that everything is always correctly priced and we will admit the possibility of occasional arbitrage opportunities. For example, it may be more profitable to hedge a portfolio with a derivative to produce a risk-free rate than to invest in the risk-free asset. In addition, we might find that replication can have lower transaction costs. For example, a derivative on a stock index combined with the risk-free asset [Long derivative (Stock index futures) + Long risk-free asset (Lending) = Long asset (Stock index)] can potentially replicate an index fund at lower transaction costs than buying all the securities in the index. Replication is the essence of arbitrage. The ability to replicate something with something else can be valuable to investors, either through pricing differentials, however temporary, or lower transaction costs.

3.4 Risk Aversion, Risk Neutrality, and Arbitrage-Free Pricing

Most investors are risk averse. They do not accept risk without the expectation of a return commensurate with that risk. Thus, they require risk premiums to justify the risk. One might think that this point implies a method for pricing derivatives based on the application of a risk premium to the expected payoff of the derivative and its risk. As we will describe later, this methodology is not appropriate in the pricing of derivatives.

As previously described, a derivative can be combined with an asset to produce a risk-free position. This fact does not mean that one *should* create such a combination. It merely means that one *can* do so. The derivative price is the price that guarantees the risk-free combination of the derivative and the underlying produces a risk-free return. The derivative price can then be inferred from the characteristics of the underlying, the characteristics of the derivative, and the risk-free rate. The investor's risk aversion is not a factor in determining the derivative price. Because the risk aversion of the investor is not relevant to pricing the derivative, one can just as easily obtain the derivative price by assuming that the investor is risk neutral. That means that the expected payoff of the derivative can be discounted at the risk-free rate rather than the risk-free rate plus a risk premium. Virtually all derivative pricing models ultimately take this form: discounting the expected payoff of the derivative at the risk-free rate.

The entire process of pricing derivatives is not exactly as we have described it at this point. There is an intermediate step, which entails altering the probabilities of the outcomes from the true probabilities to something called risk-neutral probabilities. We will illustrate this process later in this reading. The important point to understand is that while the risk aversion of investors is relevant to pricing assets, it is not relevant to pricing derivatives. As such, derivatives pricing is sometimes called **risk-neutral pricing**. Risk-neutral pricing uses the fact that arbitrage opportunities guarantee that a risk-free portfolio consisting of the underlying and the derivative must earn the risk-free rate. There is only one derivative price that meets that condition. Any mispricing of the derivative will lead to arbitrage transactions that drive the derivative price back to where it should be, the price that eliminates arbitrage opportunities.

The overall process of pricing derivatives by arbitrage and risk neutrality is called **arbitrage-free pricing**. We are effectively determining the price of a derivative by assuming that the market is free of arbitrage opportunities. This notion is also sometimes called the **principle of no arbitrage**. If there are no arbitrage opportunities, combinations of assets and/or derivatives that produce the same results must sell for the same price. The correct derivative price assures us that the market is free of arbitrage opportunities.

3.5 Limits to Arbitrage

As we previously described, there may be reasons to not pick up a coin lying on the ground. Likewise, some small arbitrage profits are never exploited. A bond selling for €1,000 might offer an arbitrage profit by trading a derivative on the bond and a risk-free asset at a total cost of €999, but the profit of €1 might be exceeded by the transaction costs. Such small differentials can easily remain essentially trapped within the bounds of transaction costs. In addition, arbitrage can require capital. Not everyone can borrow virtually unlimited amounts of money at what amounts to a risk-free rate. Moreover, some transactions can require additional capital to maintain positions. The corresponding gains from an offsetting position might not be liquid. Hence, on paper the position is hedged, but in practice, one position has a cash outflow while the other generates gains on paper that are realized only later. Borrowing against those future gains is not always easy.

Moreover, some apparent arbitrage transactions are not completely risk free. As you will learn later, option pricing requires knowledge of the volatility of the underlying asset, which is information that is not easy to obtain and subject to different opinions. Executing an arbitrage can entail risk if one lacks accurate information on the model inputs.

Some arbitrage positions require short-selling assets that can be difficult to short. Some securities are held only by investors who are unwilling to lend the securities and who, by policy, are not arbitrageurs themselves. Some commodities, in particular, can be difficult and costly to sell short. Hence, the arbitrage might exist in only one direction, which keeps the price from becoming seemingly too high or seemingly too low but permitting it to move virtually without limit in the opposite direction.

Arbitrage positions rely on the ultimate realization by other investors of the existence of the mispricing. For some investors, bearing these costs and risks until other investors drive the price back to its appropriate level can be nearly impossible.

The arbitrage principle is the essence of derivative pricing models. Yet, clearly there are limits to the ability of all investors to execute arbitrage transactions. In studying derivative pricing, it is important to accept the no-arbitrage rule as a paradigm, meaning a framework for analysis and understanding. Although no market experts think that arbitrage opportunities never occur, it is a common belief that finding and exploiting them is a challenging and highly competitive process that will not yield frequent success. But it is important that market participants stay alert for and exploit whatever arbitrage opportunities arise. In response, the market functions more efficiently.

EXAMPLE 2

Arbitrage

1 Which of the following *best* describes an arbitrage opportunity? It is an opportunity to:
   - A earn a risk premium in the short run.
   - B buy an asset at less than its fundamental value.
   - C make a profit at no risk with no capital invested.

2 What *most likely* happens when an arbitrage opportunity exists?
   - A Investors trade quickly and prices adjust to eliminate the opportunity.
   - B Risk premiums increase to compensate traders for the additional risk.
   - C Markets cease operations to eliminate the possibility of profit at no risk.

3 Which of the following *best* describes how derivatives are priced?
   - A A hedge portfolio is used that eliminates arbitrage opportunities.
   - B The payoff of the underlying is adjusted downward by the derivative value.
   - C The expected future payoff of the derivative is discounted at the risk-free rate plus a risk premium.

4 An investor who requires no premium to compensate for the assumption of risk is said to be which of the following?
   - A Risk seeking
   - B Risk averse
   - C Risk neutral

5 Which of the following is a limit to arbitrage?
   - A Clearinghouses restrict the transactions that can be arbitraged.
   - B Pricing models do not show whether to buy or sell the derivative.
   - C It may not always be possible to raise sufficient capital to engage in arbitrage.

Solution to 1:

C is correct because it is the only answer that is based on the notion of when an arbitrage opportunity exists: when two identical assets or portfolios sell for different prices. A risk premium earned in the short run can easily have occurred through luck. Buying an asset at less than fair value might not even produce a profit.

Solution to 2:

A is correct. The combined actions of traders push prices back in line to a level at which no arbitrage opportunities exist. Markets certainly do not shut down, and risk premiums do not adjust and, in fact, have no relevance to arbitrage profits.

Solution to 3:

A is correct. A hedge portfolio is formed that eliminates arbitrage opportunities and implies a unique price for the derivative. The other answers are incorrect because the underlying payoff is not adjusted by the derivative value and the discount rate of the derivative does not include a risk premium.

Solution to 4:

C is correct. Risk-seeking investors give away a risk premium because they enjoy taking risk. Risk-averse investors expect a risk premium to compensate for the risk. Risk-neutral investors neither give nor receive a risk premium because they have no feelings about risk.

Solution to 5:

C is correct. It may not always be possible to raise sufficient capital to engage in arbitrage. Clearinghouses do not restrict arbitrage. Pricing models show what the price of the derivative should be.

Thus, comparison with the market price will indicate if the derivative is overpriced and should be sold or if it is underpriced and should be purchased.

4 PRICING AND VALUATION OF FORWARD CONTRACTS: PRICING VS. VALUATION; EXPIRATION; INITIATION

b explain the difference between value and price of forward and futures contracts;

c calculate a forward price of an asset with zero, positive, or negative net cost of carry;

d explain how the value and price of a forward contract are determined at expiration, during the life of the contract, and at initiation;

e describe monetary and nonmonetary benefits and costs associated with holding the underlying asset and explain how they affect the value and price of a forward contract;

i explain the difference between value and price of swaps;

In equity markets, analysis is undertaken with the objective of determining the value, sometimes called the fundamental value, of a stock. When a stock trades in the market for a price that differs from its fundamental value, investors will often buy or sell the stock based on the perceived mispricing. The fundamental value of a stock is typically determined by analyzing the company's financial statements, projecting its earnings and dividends, determining a discount rate based on the risk, and finding the present value of the future dividends. These steps make up the essence of dividend discount models. Other approaches include comparing the book value of a company to its market value, thereby using book value as a proxy for fundamental value, or by application of a price/earnings ratio to projected next-period earnings, or by discounting free cash flow. Each of these approaches purports to estimate the company's fundamental value, leading to the notion that a company is worth something that may or may not correspond to its price in the market.

In derivative markets, the notion of valuation as a representation of fundamental value is still a valid concept, but the terminology can be somewhat different and can lead to some confusion. Options are not a problem in this regard. They can be analyzed to determine their fundamental value, and the market price can be compared with the fundamental value. Any difference can then presumably be exploited via arbitrage. The combined actions of numerous investors should ultimately lead to the market price converging to its fundamental value, subject to the above limits to arbitrage.

The world of forwards, futures, and swaps, however, uses different terminology with respect to price and value. These contracts do not require the outlay of cash at the start the way an option, stock, or bond does. Forwards, futures, and swaps start off with values of zero. Then as the underlying moves, their values become either positive or negative. The forward, futures, or swap price is a concept that represents the fixed price or rate at which the underlying will be purchased at a later date. It is not an amount to be paid at the start. This fixed price or rate is embedded into the contract while the value will fluctuate as market conditions change. But more importantly, the value and price are not at all comparable with each other.

Consider a simple example. Suppose you own a stock priced at \$102. You have a short forward contract to sell the stock at a price of \$100 one year from now. The risk-free rate is 4%. Your position is riskless because you know that one year from now, you will sell the stock for \$100. Thus, you know you will get \$100 one year from now, which has a present value of \$100/(1.04) = \$96.15. Notice the discounting at the risk-free rate, which is appropriate because the position is riskless. Your overall position is that you own an asset worth \$102 and are short a contract worth something, and the two positions combine to have a value of \$96.15. Therefore, the forward contract must have a value of \$96.15 – \$102 = –\$5.85. Your forward contract is thus worth –\$5.85. To the party on the opposite side, it is worth +\$5.85.[7] The price of the forward contract is still \$100, which was set when you created the contract at an earlier date. As you can see, the \$100 forward price is not comparable to the \$5.85 value of the contract.

Although the forward price is fixed, any new forward contract calling for delivery of the same asset at the same time will have a different price. We will cover that point in more detail later. For now, it is important to see that your contract has a price of \$100 but a value of –\$5.85, which are two entirely different orders of magnitude. This information does not imply that the forward contract is mispriced. The value is the amount of wealth represented by owning the forward contract. The price is one of the terms the parties agreed on when they created the contract.[8] This idea applies in the same manner for futures and swaps.

4.1 Pricing and Valuation of Forward Commitments

In this section, we will go into pricing forward commitments in a little more detail. Let us start by establishing that today, at time 0, we create a forward commitment that expires at time T. The value of the underlying today is S_0. At expiration the underlying value is S_T, which is not known at the initiation of the contract.

4.1.1 *Pricing and Valuation of Forward Contracts*

Previously, we noted that price and value are entirely different concepts for forward commitments. We gave an example of a forward contract with a price of \$100 but a value of –\$5.85 to the seller and +\$5.85 to the buyer. In the next subsection, we will delve more deeply into understanding these concepts of pricing and valuation for forward contracts.

7 This concept of the value of the forward contract as it evolves toward expiration is sometimes referred to as its mark-to-market value. The same notion is applicable to swaps. In futures, of course, contracts are automatically marked to market by the clearinghouse, and gains and losses are converted into actual cash flows from one party to the other.

8 The forward price is more like the exercise price of the option. It is the price the two parties agree will be paid at a future date for the underlying. Of course, the option has the feature that the holder need not ever pay that price, which is the case if the holder chooses not to exercise the option.

4.1.1.1 Pricing and Valuation of Forward Contracts at Expiration Recall that a forward contract specifies that one party agrees to buy the underlying from the other at the expiration date at a price agreed on at the start of the contract. Suppose that you enter into a contract with another party in which you will buy a used car from that party in one year at a price of $10,000. Then $10,000 is the forward price. One year later, when the contract expires, you are committed to paying $10,000 and accepting delivery of the car. Let us say that at that time, you check the used car market and find that an identical car is worth $10,800. How much is your forward contract worth to you at that time? It obligates you to pay $10,000 for a car that you would otherwise have to pay $10,800. Thus, the contract benefits you by $800, so its value is $800. If you were on the opposite side of the transaction, its value would be –$800. If the market price of the car were below $10,000, the contract would have negative value to you and the mirror image positive value to the seller.

This example leads us to our first important derivative pricing result. The forward price, established at the initiation date of contract is $F_0(T)$. Let us denote the value at expiration of the forward contract as $V_T(T)$. This value is formally stated as

$$V_T(T) = S_T - F_0(T) \tag{1}$$

In words,

> *The value of a forward contract at expiration is the spot price of the underlying minus the forward price agreed to in the contract.*

In the financial world, we generally define value as the value to the long position, so the above definition is generally correct but would be adjusted if we look at the transaction from the point of view of the short party. In that case, we would multiply the value to the long party by −1 to calculate the value to the short party. Alternatively, the value to the short party is the forward price minus the spot price at expiration.

If a forward contract could be initiated right at the instant of expiration, the forward price would clearly be the spot price. Such a contract would essentially be a spot transaction.

4.1.1.2 Pricing and Valuation at Initiation Date In Exhibit 7, we see the nature of the problem of pricing a forward contract. We are situated at time 0, facing an uncertain future. At the horizon date, time T, the underlying price will be S_T. Of course, at time 0 we do not know what S_T will turn out to be. Yet at time 0, we need to establish the forward price, $F_0(T)$, which is the price we agree to pay at time T to purchase the asset.

Exhibit 7 The Time Horizon of Forward Contracts

When a forward contract is initiated, neither party pays anything to the other. It is a valueless contract, neither an asset nor a liability. Therefore, its value at initiation is zero:

$$V_0(T) = 0 \tag{2}$$

The forward price that the parties agree to at the initiation date of the contract is a special price that results in the contract having zero value and prohibiting arbitrage. This is our first important result:

> *Because neither the long nor the short pays anything to the other at the initiation date of a forward contract, the value of a forward contract when initiated is zero.*

If this statement were not true and one party paid a sum of money to the other, the party receiving the money could find another party and engage in the opposite transaction, with no money paid to the other on this second contract. The two transactions would completely offset, thereby eliminating the risk. Yet, the first party would have captured some cash from the second and consequently earned an arbitrage profit because his position is completely hedged. He would walk away with money and never have to worry about paying it back. The forward price is the price the two parties agree on that generates a value of zero at the initiation date. Finding that price is actually quite easy.

Consider a very simple asset price at S_0 today that pays no dividends or interest, nor does it yield any nonfinancial benefits or incur any carrying costs. As described earlier, we can peer into the future, but at best we can make only a forecast of the price of this asset at our horizon date of time T. That forecast was previously referred to as the expected spot price at expiration, $E(S_T)$. On the surface, it might seem that pricing a forward contract would somehow involve a discounting of the expected spot price. As we said earlier, however, that is not how derivatives are priced—they are priced using arbitrage.

Suppose we hold the asset and enter into a forward contract to sell the asset at the price $F_0(T)$. It should be easy to see that we have constructed a risk-free position. We know that the asset, currently worth S_0, will be sold later at $F_0(T)$ and that this price should guarantee a risk-free return. Thus, we should find the following relationship,

$$\frac{F_0(T)}{S_0} = (1 + r)^T \tag{3}$$

We can easily solve for the forward price to obtain

$$F_0(T) = S_0(1 + r)^T \tag{4}$$

Or, in words,

> *The forward price is the spot price compounded at the risk-free rate over the life of the contract.*

There is a nice logic to this relationship. While the spot price is what someone would have to pay today to buy the asset, a forward contract locks in the purchase price at the horizon date. When that date arrives, the investor will own the asset. Instead of buying the asset today, suppose the investor uses the forward contract to guarantee that she will own the asset at the horizon date. By using the forward contract, the investor will not have committed the money, S_0, that would have forgone interest at the rate r for the period 0 to T. Notice how the risk premium on the asset does not directly appear in the pricing relationship. It does appear implicitly, because it determines the spot price paid to buy the asset. Knowing the spot price, however, eliminates the necessity of determining the risk premium. The derivatives market can simply let the spot market derive the risk premium.

As a simple example, let us say the underlying price, S_0, is £50, the risk-free rate, r, is 3%, and the contract expires in three months, meaning that $T = 3/12 = 0.25$. Then the forward price is £50(1.03)$^{0.25}$ = £50.37. Thus, the two parties would agree that the buyer will pay £50.37 to the seller in three months, and the seller will deliver the underlying to the buyer at expiration.

Now suppose the asset generates cash payments and/or benefits and incurs storage costs. As we discussed, the net cost of carry consists of the benefits, denoted as γ (dividends or interest plus convenience yield), minus the costs, denoted as θ, both of which are in present value form. To put these concepts in future value form, we simply compound them at the risk-free rate, $(\gamma - \theta)(1 + r)^T$. Because this is their value at the expiration date of the contract, we can add them to $F_0(T)$ in Equation 3, thereby restating that equation as

$$(1 + r)^T = \frac{F_0(T) + (\gamma - \theta)(1 + r)^T}{S_0}$$

The numerator is how much money we end up with at T. Rearranging, we obtain the forward price as

$$F_0(T) = (S_0 - \gamma + \theta)(1 + r)^T$$

or **(5)**

$$F_0(T) = S_0(1 + r)^T - (\gamma - \theta)(1 + r)^T$$

From Equation 5, we can see that the forward price determined using Equation 4 is reduced by the future value of any benefits and increased by the future value of any costs. In other words,

> *The forward price of an asset with benefits and/or costs is the spot price compounded at the risk-free rate over the life of the contract minus the future value of those benefits and costs.*

Again, the logic is straightforward. To acquire a position in the asset at time T, an investor could buy the asset today and hold it until time T. Alternatively, he could enter into a forward contract, committing him to buying the asset at T at the price $F_0(T)$. He would end up at T holding the asset, but the spot transaction would yield benefits and incur costs, whereas the forward transaction would forgo the benefits but avoid the costs.

Assume the benefits exceed the costs. Then the forward transaction would return less than the spot transaction. The formula adjusts the forward price downward by the expression $-(\gamma - \theta)(1 + r)^T$ to reflect this net loss over the spot transaction. In other words, acquiring the asset in the forward market would be cheaper because it forgoes benefits that exceed the costs. That does not mean the forward strategy is better. It costs less but also produces less. Alternatively, if the costs exceeded the benefits, the forward price would be higher because the forward contract avoids the costs at the expense of the lesser benefits.

Returning to our simple example, suppose the present value of the benefits is γ = £3 and the present value of the costs is θ = £4. The forward price would be £50$(1.03)^{0.25}$ – (£3 – £4)$(1.03)^{0.25}$ = £51.38. The forward price, which was £50.37 without these costs and benefits, is now higher because the carrying costs exceed the benefits.

The value of the contract when initiated is zero provided the forward price conforms to the appropriate pricing formula. To keep the analysis as simple as possible, consider the case in which the asset yields no benefits and incurs no costs. Going long the forward contract or going long the asset produces the same position at T: ownership of the asset. Nonetheless, the strategies are not equivalent. Going long the forward contract enables the investor to avoid having to pay the price of the asset, S_0, so she would collect interest on the money. Thus, the forward strategy would have a value of S_0, reflecting the investment of that much cash invested in risk-free bonds, plus the value of the forward contract. The spot strategy would have a value of S_0, reflecting the investment in the asset. These two strategies must have equal values. Hence, the value of the forward contract must be zero.

Although a forward contract has zero value at the start, it will not have zero value during its life. We now take a look at what happens during the life of the contract.

PRICING AND VALUATION OF FORWARD CONTRACTS: BETWEEN INITIATION AND EXPIRATION; FORWARD RATE AGREEMENTS

5

b explain the difference between value and price of forward and futures contracts;

c calculate a forward price of an asset with zero, positive, or negative net cost of carry;

d explain how the value and price of a forward contract are determined at expiration, during the life of the contract, and at initiation;

e describe monetary and nonmonetary benefits and costs associated with holding the underlying asset and explain how they affect the value and price of a forward contract;

f define a forward rate agreement and describe its uses;

We previously worked an example in which a forward contract established with a price of \$100 later has a value of –\$5.85 to the seller and +\$5.85 to the buyer. Generally we would say the value is \$5.85. We explained that with the spot price at \$102, a party that is long the asset and short the forward contract would guarantee the sale of the asset priced at \$102 at a price of \$100 in one year. The present value of \$100 in one year at 4% is \$96.15. Thus, the party guarantees that his \$102 asset will be effectively sold at a present value of \$96.15, for a present value loss of \$5.85.

In general, we can say that

> *The value of a forward contract is the spot price of the underlying asset minus the present value of the forward price.*

Again, the logic is simple. A forward contract provides a type of synthetic position in the asset, for which we promise to pay the forward price at expiration. Thus, the value of the forward contract is the spot price of the asset minus the present value of the forward price. Let us write out this relationship using $V_t(T)$ as the value of the forward contract at time t, which is some point in time after the contract is initiated and before it expires:

$$V_t(T) = S_t - F_0(T)(1 + r)^{-(T-t)} \tag{6}$$

Note that we are working with the spot price at t, but the forward price was fixed when the contract was initiated.[9]

Now, recall the problem we worked in which the underlying had a price of £50 and the contract was initiated with a three-month life at a price of £50.37. Move one month later, so that the remaining time is two months: $T - t = 2/12 = 0.167$. Let the underlying price be £52. The value of the contract would be £52 – £50.37$(1.03)^{-0.167}$ = £1.88.

If the asset has a cost of carry, we must make only a small adjustment:

$$V_t(T) = S_t - (\gamma - \theta)(1 + r)^t - F_0(T)(1 + r)^{-(T-t)} \tag{7}$$

Note how we adjust the formula by the net of benefits minus costs. The forward contract forgoes the benefits and avoids the costs of holding the asset. Consequently, we adjust the value downward to reflect the forgone benefits and upward to reflect the avoided costs. Remember that the costs (θ) and benefits (γ) are expressed on a present

9 An alternative approach to valuing a forward contract during its life is to determine the price of a new forward contract that would offset the old one. The discounted difference between the new forward price and the original forward price will lead to the same value.

value basis as of time 0. We need their value at time t. We could compound them from 0 to T and then discount them back to t by the period $T - t$, but a shorter route is to simply compound them from 0 to t. In the problem we previously worked, in which we priced the forward contract when the asset has costs and benefits, the benefits (γ) were £3 and the costs (θ) were £4, giving us a forward price of £51.38. We have now moved one month ahead, so $t = 1/12 = 0.0833$ and $T - t = 2/12 = 0.167$. Hence the value of the forward contract would be £52 – (£3 – £4)$(1.03)^{0.0833}$ – £51.38$(1.03)^{-0.167}$ = £1.88. Notice how the answer is the same as in the case of no costs and benefits, as this effect is also embedded in the original forward price and completely offsets.

It is important to note that although we say that Equation 7 holds during the life of the contract at some arbitrary time t, it also holds at the initiation date and at expiration. For the initiation date, we simply change t to 0 in Equation 7. Then we substitute Equation 5 for $F_0(T)$ in Equation 7, obtaining $V_0(T) = 0$, confirming that the value of a forward contract at initiation is zero. At expiration, we let $t = T$ in Equation 7 and obtain the spot price minus the forward price, as presented in Equation 1.[10]

5.1 A Word about Forward Contracts on Interest Rates

Forward contracts in which the underlying is an interest rate are called **forward rate agreements**, or FRAs. These instruments differ slightly from most other forward contracts in that the underlying is not an asset. Changes in interest rates, such as the value of an asset, are unpredictable. Moreover, virtually every company and organization is affected by the uncertainty of interest rates. Hence, FRAs are very useful devices for many companies. FRAs are forward contracts that allow participants to make a known interest payment at a later date and receive in return an unknown interest payment. In that way, a participant whose business will involve borrowing at a future date can hedge against an increase in interest rates by buying an FRA (the long side) and locking in a fixed payment and receiving a random payment that offsets the unknown interest payment it will make on its loan. Note that the FRA seller (the short side) is hedging against a decrease in interest rates. Also, consider that the FRA seller could be a lender wishing to lock in a fixed rate on a loan it will make at a future date.

Even though FRAs do not involve an underlying asset, they can still be combined with an underlying asset to produce a hedged position, thereby leading to fairly straightforward pricing and valuation equations. The math is a little more complex than the math for forwards on assets, but the basic ideas are the same.

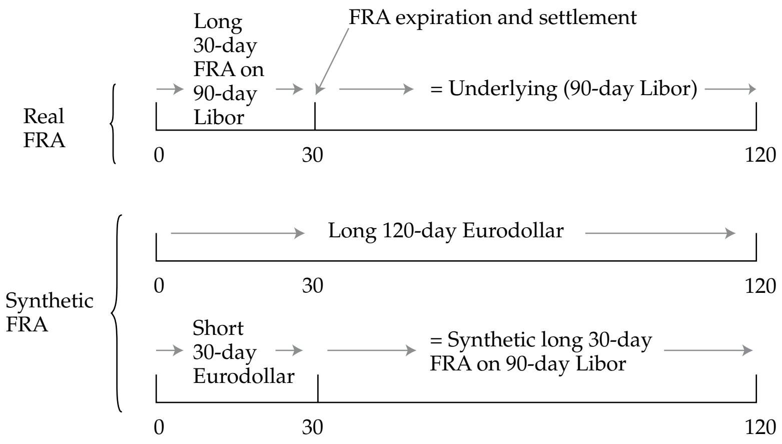
FRAs have often historically been based on Libor, the London Interbank Offered Rate, which represents the rate on a Eurodollar time deposit, a loan in dollars from one London bank to another. Other rates such as Euribor (Euro Interbank Offered Rate) and Tibor (Tokyo Interbank Offered Rate) have also been used.[11] As an example, assume we are interested in going long a 30-day FRA with a fixed rate (the FRA rate) in which the underlying is 90-day Libor. A long position means that in 30 days, we will make a known interest payment and receive an interest payment corresponding to the discounted difference between 90-day Libor on that day and the FRA rate. We can either enter into a 30-day FRA on 90-day Libor or create a synthetic FRA. To do the latter, we would go long a 120-day Eurodollar time deposit and short a 30-day Eurodollar time deposit. Exhibit 8 shows the structure of this strategy. We omit some of the details here, such as how much face value we should take on the two Eurodollar transactions as well as the size of the FRA. Those technical issues are covered in

10 You might be wondering whether the cost and benefit terms disappear when $t = T$. With the costs and benefits defined as those incurred over the period t to T, at expiration their value is zero by definition.

11 Libor is being phased out, as the panel of banks will no longer be required to submit quotations after 2021. In anticipation of this, market participants and regulators have been working to develop alternative reference rates.

more advanced material. At this time, we focus on the fact that going long over the 120-day period and short over the 30-day period leaves an investor with no exposure over the 30-day period and then converts to a position that starts 30 days from now and matures 90 days later. This synthetic position corresponds to a 30-day FRA on 90-day Libor. Exhibit 8 illustrates this point.[12]

Exhibit 8 Real FRA and Synthetic FRA (30-Day FRA on 90-Day Libor)



FRAs, and indeed all forward contracts relating to bonds and interest rates, are closely tied to the term structure of interest rates, a concept covered in virtually all treatments of fixed-income securities. Buying a 120-day zero-coupon bond and selling a 30-day zero-coupon bond produces a forward position in a 90-day zero-coupon bond that begins in 30 days. From that forward position, one can infer the forward rate. It would then be seen that the FRA rate *is* the forward rate, even though the derivative itself is not a forward contract on a bond.

EXAMPLE 3

Forward Contract Pricing and Valuation

1 Which of the following *best* describes the difference between the price of a forward contract and its value?

   A The forward price is fixed at the start, and the value starts at zero and then changes.

   B The price determines the profit to the buyer, and the value determines the profit to the seller.

   C The forward contract value is a benchmark against which the price is compared for the purposes of determining whether a trade is advisable.

12 The real FRA we show appears to imply that an investor enters into a Eurodollar transaction in 30 days that matures 90 days later. This is not technically true. The investor does, however, engage in a cash settlement in 30 days that has the same value and economic form as such a transaction. Specifically, settlement at expiration of the FRA is an amount equal to the discounted difference between the underlying 90-day Libor rate on that day and the FRA rate multiplied by a notional principal amount. These details are covered in the Level II and Level III CFA Program curriculum.

2 Which of the following *best* describes the value of the forward contract at expiration? The value is the price of the underlying:
   A minus the forward price.
   B divided by the forward price.
   C minus the compounded forward price.
3 Which of the following factors does *not* affect the forward price?
   A The costs of holding the underlying
   B Dividends or interest paid by the underlying
   C Whether the investor is risk averse, risk seeking, or risk neutral
4 Which of the following *best* describes the forward rate of an FRA?
   A The spot rate implied by the term structure
   B The forward rate implied by the term structure
   C The rate on a zero-coupon bond of maturity equal to that of the forward contract

Solution to 1:

A is correct. The forward price is fixed at the start, whereas the value starts at zero and then changes. Both price and value are relevant in determining the profit for both parties. The forward contract value is not a benchmark for comparison with the price.

Solution to 2:

A is correct because the holder of the contract gains the difference between the price of the underlying and the forward price. That value can, of course, be negative, which will occur if the holder is forced to buy the underlying at a price higher than the market price.

Solution to 3:

C is correct. The costs of holding the underlying, known as carrying costs, and the dividends and interest paid by the underlying are extremely relevant to the forward price. How the investor feels about risk is irrelevant, because the forward price is determined by arbitrage.

Solution to 4:

B is correct. FRAs are based on Libor, and they represent forward rates, not spot rates. Spot rates are needed to determine forward rates, but they are not equal to forward rates. The rate on a zero-coupon bond of maturity equal to that of the forward contract describes a spot rate.

As noted, we are not covering the details of derivative pricing but rather are focusing on the intuition. At this point, we have covered the intuition of pricing forward contracts. We now move to futures contracts.

6 PRICING AND VALUATION OF FUTURES CONTRACTS

b explain the difference between value and price of forward and futures contracts;
g explain why forward and futures prices differ;

Futures contracts differ from forward contracts in that they have standard terms, are traded on a futures exchange, and are more heavily regulated, whereas forward contracts are typically private, customized transactions. Perhaps the most important distinction is that they are marked to market on a daily basis, meaning that the accumulated gains and losses from the previous day's trading session are deducted from the accounts of those holding losing positions and transferred to the accounts of those holding winning positions. This daily settling of gains and losses enables the futures exchange to guarantee that a party that earns a profit from a futures transaction will not have to worry about collecting the money. Thus, futures exchanges provide a credit guarantee, which is facilitated through the use of a clearinghouse. The clearinghouse collects and disburses cash flows from the parties on a daily basis, thereby settling obligations quickly before they accumulate to much larger amounts. There is no absolute assurance that a clearinghouse will not fail, but none has ever done so since the first one was created in the 1920s.

The pattern of cash flows in a futures contract is quite similar to that in a forward contract. Suppose you enter into a forward contract two days before expiration in which you agree to buy an asset at €100, the forward price. Two days later, the asset is selling for €103, and the contract expires. You therefore pay €100 and receive an asset worth €103, for a gain of €3. If the contract were cash settled, instead of involving physical delivery, you would receive €3 in cash, which you could use to defer a portion of the cost of the asset. The net effect is that you are buying the asset for €103, paying €100 plus the €3 profit on the forward contract.

Had you chosen a futures contract, the futures price at expiration would still converge to the spot price of €103. But now it would matter what the futures settlement price was on the next to last day. Let us assume that price was €99. That means on the next to last day, your account would be marked to market for a loss of €1, the price of €100 having fallen to €99. That is, you would be charged €1, with the money passed on to the opposite party. But then on the last day, your position would be marked from €99 to €103, a gain of €4. Your net would be €1 lost on the first day and €4 gained on the second for a total of €3. In both situations you gain €3, but with the forward contract, you gain it all at expiration, whereas with the futures contract, you gain it over two days. With this two-day example, the interest on the interim cash flow would be virtually irrelevant, but over longer periods and with sufficiently high interest rates, the difference in the amount of money you end up with could be noticeable.

The value of a futures contract is the accumulated gain or loss on a futures contract since its previous day's settlement. When that value is paid out in the daily settlement, the futures price is effectively reset to the settlement price and the value goes to zero. The different patterns of cash flows for forwards and futures can lead to differences in the pricing of forwards versus futures. But there are some conditions under which the pricing is the same. It turns out that if interest rates were constant, forwards and futures would have the same prices. The differential will vary with the volatility of interest rates. In addition, if futures prices and interest rates are uncorrelated, forwards and futures prices will be the same. If futures prices are positively correlated with interest rates, futures contracts are more desirable to holders of long positions than are forwards. The reason is because rising prices lead to futures profits that are reinvested in periods of rising interest rates, and falling prices leads to losses that occur in periods of falling interest rates. It is far better to receive cash flows in the interim than all at expiration under such conditions. This condition makes futures more attractive than forwards, and therefore their prices will be higher than forward prices. A negative correlation between futures prices and interest rates leads to the opposite interpretation, with forwards being more desirable than futures to the long position. The more desirable contract will tend to have the higher price.

The practical realities, however, are that the derivatives industry makes virtually no distinction between futures and forward prices.[13] Thus, we will make no distinction between futures and forward pricing, except possibly in noting some subtle issues that may arise from time to time.

EXAMPLE 4

Futures Pricing and Valuation

1 Which of the following *best* describes how futures contract payoffs differ from forward contract payoffs?
   - A Forward contract payoffs are larger.
   - B They are equal, ignoring the time value of money.
   - C Futures contract payoffs are larger if the underlying is a commodity.

2 Which of the following conditions will not make futures and forward prices equivalent?
   - A Interest rates are constant.
   - B Futures prices are uncorrelated with interest rates.
   - C The volatility of the forward price is different from the volatility of the futures price.

3 With respect to the value of a futures contract, which of the following statements is *most* accurate? The value is the:
   - A futures price minus the spot price.
   - B present value of the expected payoff at expiration.
   - C accumulated gain since the previous settlement, which resets to zero upon settlement.

Solution to 1:

B is correct. Forward payoffs occur all at expiration, whereas futures payoffs occur on a day-to-day basis but would equal forward payoffs ignoring interest. Payoffs could differ, so forward payoffs are not always larger. The type of underlying is not relevant to the point of which payoff is larger.

Solution to 2:

C is correct. Constant interest rates or the condition that futures prices are uncorrelated with interest rates will make forward and futures prices equivalent. The volatility of forward and futures prices has no relationship to any difference.

Solution to 3:

C is correct. Value accumulates from the previous settlement and goes to zero when distributed.

13 At the time of this writing, many forwards (and swaps) are being processed through clearinghouses, a response to changes brought about by key legislation in several countries that was adopted following the financial crises of 2008. These OTC instruments are thus being effectively marked to market in a similar manner to the futures contracts described here. The full extent of this evolution of OTC trading through clearinghouses is not yet clear.

PRICING AND VALUATION OF SWAP CONTRACTS

7

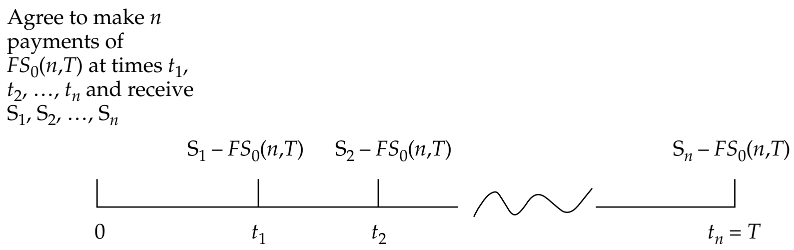
h explain how swap contracts are similar to but different from a series of forward contracts;

i explain the difference between value and price of swaps;

Recall the structure of a forward contract, as depicted in Exhibit 7. The investor is at time 0 and needs to determine the price, $F_0(T)$, that she will agree to pay at time T to purchase the asset. This price is set such that there is no value to the contract at that time. Value can arise later as prices change, but when initiated, the contract has zero value. Neither party pays anything to the other at the start.

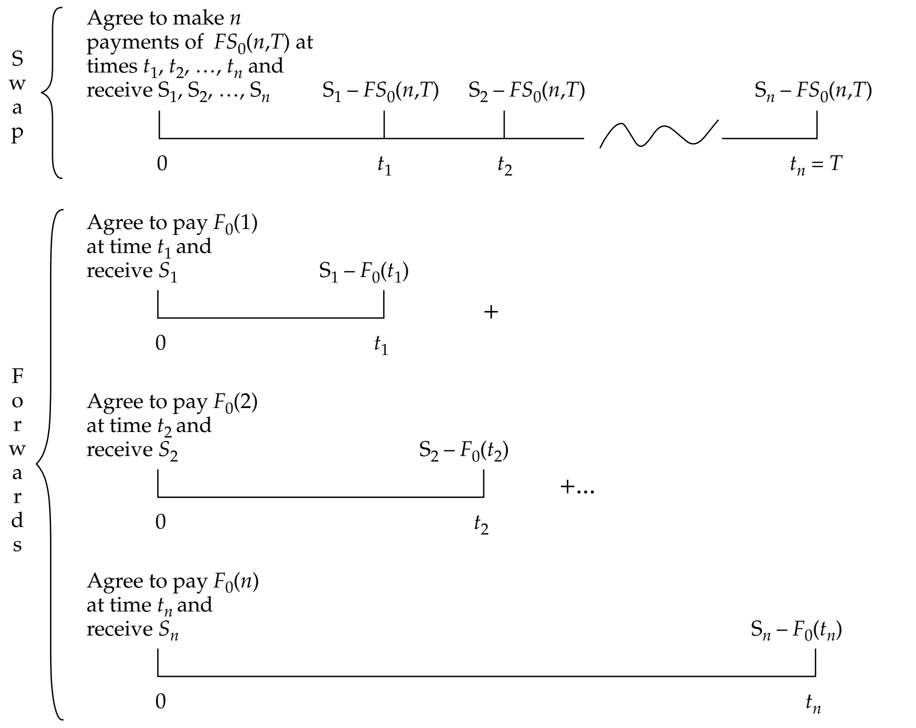
Now consider a swap starting at time 0 and ending at time T. We will let this swap be the type that involves a fixed payment exchanged for a floating payment. The contract specifies that the two parties will make a series of n payments at times that we will designate as 1, 2, …, n, with the last payment occurring at time T. On each of these payment dates, the owner of the swap makes a payment of $FS_0(n, T)$ and receives a payment based on the value of the underlying at the time of each respective payment, $S_1, S_2, \ldots, S_n$. So from the point of view of the buyer, the sequence of cash flows from the swap is $S_1 - FS_0(n, T)$, $S_2 - FS_0(n, T)$, …, $S_n - FS_0(n, T)$. The notation $FS_0(n, T)$ denotes the fixed payment established at time 0 for a swap consisting of n payments with the last payment at time T. We denote the time to each payment as $t_1, t_2, \ldots, t_n$, where $t_n = T$. This structure is shown in Exhibit 9.

Exhibit 9 Structure of Cash Flows in a Swap



Comparing Exhibit 7 with Exhibit 9 reveals some similarities. A swap is in some sense a series of forward contracts, specifically a set of contracts expiring at various times in which one party agrees to make a fixed payment and receive a variable payment. Now consider Exhibit 10, which breaks down a swap into a series of implicit forward contracts, with the expiration of each forward contract corresponding to a swap payment date.

Exhibit 10 A Swap as a Series of Forward Contracts



Recall from the material on forward contracts that the forward price is determined by the spot price and the net cost of carry (Equation 5), the latter being partially determined by the length of time of the contract. It should be obvious that a forward contract expiring at time t_1 will not have the same price, $F_0(t_1)$, as a forward contract expiring at time t_2, $F_0(t_2)$, and likewise for all of the implicit remaining forward contracts expiring up through time t_n. The cost of carrying an asset over different time periods will vary by the length of the time periods. In other words, the prices of the implicit forward contracts embedded in a swap will not be equal:

$$F_0(\mathrm{t}_1) \neq F_0(t_2) \neq \ldots \neq F_0(t_n)$$

But for a swap, all the fixed payments are equal. So, how can we equate a swap to a series of forward contracts? It turns out that we can, and in doing so, we recall a valuable point about forward pricing.

Recall that the forward price is the price that produces a zero value of the contract at the start. Zero value is essential if there is no exchange of cash flows from one party to the other. And although no exchange of cash flows is customary, it is not mandatory. The parties could agree on any forward price at the start. If the zero-value forward price were \$30 and the parties agreed on a price of \$28, it should be apparent that the buyer would be getting a great price. The seller, being rational, would require that the buyer compensate him at the start. The seller should be getting \$30 at expiration and instead will get \$28. So the buyer should compensate the seller to the amount of the present value of \$2 at expiration. If the parties agree on a price greater than \$30, similar compensation would have to be paid from seller to buyer.

A forward transaction that starts with a zero value is called an at-market forward. A forward transaction that starts with a nonzero value is called an off-market forward. There is generally no prohibition on the use of off-market forward contracts, so two parties can engage in a series of forward contracts at whatever fixed price they so desire. Assume they agree on the price $FS_0(T)$. That is, each forward contract will be created at the fixed price that corresponds to the fixed price of a swap of the same maturity with payments made at the same dates as the series of forward contracts. That means that some of the forward contracts would have positive values and some would have negative values, but their combined values would equal zero.

Now, it sounds like that price would be hard to find, but it is not. We would not, however, go about finding it by taking random guesses. Doing so would take seemingly forever. Along the way, we would notice that some of these implicit forward contracts would have positive values and some would have negative values. If the positives outweighed the negatives, then the overall swap value would be positive, which is too high. Likewise, we might plug in a number that would produce an overall negative value, with the implicit forward contract values tending to be predominantly negative, which is too low.

Not surprisingly, we can find that price easily by appealing to the principle of arbitrage. We said that the principle of arbitrage will guide us *all the way through* derivative pricing. We will omit the details, but here is the general idea.

Suppose we buy an asset that pays the amounts S_1, S_2, ..., S_n at times t_1, t_2, ..., t_n. These are unknown amounts. A simple example would be a floating-rate bond for which the S values represent the coupons that are unknown at the start but ultimately are determined by the evolution of interest rates. Then suppose we finance the purchase of that asset by borrowing money that we promise to repay with equal fixed payments of $FS_0(T)$. That strategy replicates the swap. As you have already learned, replication is the key to pricing.

Valuation of the swap during its life again appeals to replication and the principle of no arbitrage. We will find a way to reproduce the remaining payments on the swap with other transactions. The value of that strategy is the value of the swap.

To obtain the fixed rate on the swap or to value it later during its life, we will need information from the market for the underlying. As we previously noted, there are derivatives on bonds and interest rates, equities, currencies, and commodities. It is not possible to provide a general and simple statement of how to price swaps that covers all of these cases, but that topic is covered in advanced material.

EXAMPLE 5

Swap Pricing and Valuation

1. A swap is equivalent to a series of:
   - **A** forward contracts, each created at the swap price.
   - **B** long forward contracts, matched with short futures contracts.
   - **C** forward contracts, each created at their appropriate forward prices.
2. If the present value of the payments in a forward contract or swap is not zero, which of the following is most likely to be true?
   - **A** The contract cannot legally be created.
   - **B** The contract must be replicated by another contract with zero value.
   - **C** The party whose stream of payments to be received is greater has to pay the other party the present value difference.

Solution to 1:

A is correct. Each implicit forward contract is said to be off-market, because it is created at the swap price, not the appropriate forward price, which would be the price created in the forward market.

Solution to 2:

C is correct. The party whose stream of payments to be received is greater has to pay the other party the present value difference. Such a contract can legally be created, but the party receiving the greater present value must compensate the other party with a cash payment at the start. Replication is never required.

8 PRICING AND VALUATION OF OPTIONS

j explain the exercise value, time value, and moneyness of an option;

k identify the factors that determine the value of an option and explain how each factor affects the value of an option;

o explain under which circumstances the values of European and American options differ.

Unlike a forward, futures, or swap contract, an option is clearly an asset to the holder and a liability to the seller. The buyer of an option pays a sum of money, called the premium, and receives the right to buy (a call) or sell (a put) the underlying. The seller receives the premium and undertakes a potential obligation because the buyer has the right, but not the obligation, to exercise the option. Options are, therefore, contingent claims. Pricing the option is the same as assigning its value. Some confusion from that terminology may still arise, in that an option could trade in the market for an amount that differs from its value.

As mentioned, there are two general types of options. Calls represent the right to buy, and puts represent the right to sell. There are also two important exercise characteristics of options. American options allow exercise at any time up to the expiration, while European options allow exercise only at expiration. It is important to understand that the terms "American" and "European" have no relationship to where the options are traded. Because the right to exercise can be a complex feature of an option, European options are easier to understand, and we will focus on them first.

We will use the same notation used with forwards. We start by assuming that today is time 0, and the option expires at time T. The underlying is an asset currently priced at S_0, and at time T, its price is S_T. Of course, we do not know S_T until we get to the expiration. The option has an exercise or strike price of X. The symbols we use are as follows:

For calls,

c_0 = value (price) of European call today
c_T = value (price) of European call at expiration
C_0 = value (price) of American call today
C_T = value (price) of American call at expiration

For puts,

p_0 = value (price) of European put today
p_T = value (price) of European put at expiration
P_0 = value (price) of American put today
P_T = value (price) of American put at expiration

8.1 European Option Pricing

Recall that in studying forward contracts earlier in this reading, the first thing we learned is how a forward contract pays off at expiration. Then we backed up and determined how forward contracts are priced and valued prior to expiration. We follow that same approach for options.

8.1.1 *Value of a European Option at Expiration*

Recall that a European call option allows the holder to buy the underlying at expiration by paying the exercise price. Therefore, exercise is justified only if the value of the underlying exceeds the exercise price. Otherwise, the holder would simply let the call expire. So if the call is worth exercising ($S_T > X$), the holder pays X and receives an asset worth S_T. Thus, the option is worth $S_T - X$. If the call is not worth exercising ($S_T \leq X$), the option simply expires and is worth nothing at expiration.[14] Thus, the value of the option at expiration is the greater of either zero or the underlying price at expiration minus the exercise price, which is typically written as

$$c_T = \text{Max}(0,S_T - X) \tag{8}$$

This formula is also sometimes referred to as the **exercise value** or **intrinsic value**. In this reading, we will use the term exercise value.

Taking a simple example, if the exercise price is €40 and the underlying price is at expiration €43, the call is worth c_T = Max(€0,€43 – €40) = Max(€0,€3) = €3. If the underlying price at expiration is €39, the call is worth c_T = Max(0,€39 – €40) = Max(€0,–€1) = €0.

For puts, the holder has the right to sell the underlying at X. If the underlying is worth less than X at expiration ($X > S_T$), the put will be exercised and worth $X - S_T$ because it allowed the holder to avoid the loss in value of the asset of that amount. If the underlying is equal to or worth more than the exercise price at expiration ($S_T \geq X$), the put will simply expire with no value. So, the put is worth the greater of either zero or the exercise price minus the price of the underlying at expiration.

$$p_T = \text{Max}(0,X - S_T) \tag{9}$$

As discussed above, this formula is referred to as the exercise value or intrinsic value, and as noted, we will use the term exercise value.

Using the same example as with the call, if the underlying is €43 at expiration, the put is worth p_T = Max(€0,€40 – €43) = Max(0,–€3) = €0. If the underlying is €39 at expiration, the put is worth p_T = Max(€0,€40 – €39) = Max(€0,€1) = €1.

Thus, the holder of an option looks out into the future and sees these relationships as the payoff possibilities. That does not mean the holder knows what S_T will be, but the holder knows that all of the uncertainty of the option payoff is determined by the behavior of the underlying.

The results of this section can be restated as follows:

> *The value of a European call at expiration is the exercise value, which is the greater of zero or the value of the underlying minus the exercise price.*
>
> *The value of a European put at expiration is the exercise value, which is the greater of zero or the exercise price minus the value of the underlying.*

---

14 In all the remaining material, we identify conditions at expiration, such as $S_T > X$ and $S_T \leq X$. Here we merged the equality case ($S_T = X$) with the less-than case (<). We could have done it the other way around ($S_T < X$ and $S_T \geq X$), which would have had no effect on our interpretations or any calculations of option value. For convenience, in some situations we will use one specification and in some the other.

To understand option pricing, we have to work our way forward in a gradual manner. The next valuable steps involve using our intuition to identify some characteristics that will influence the value of the option. We might not be able to quantify their effects just yet, but we can rationalize why these factors affect the value of an option.

8.1.2 *Effect of the Value of the Underlying*

The value of the underlying is obviously a critical element in determining the value of an option. It is the uncertainty of the underlying that provides the motivation for using options. It is easy to rationalize the direction of the effect of the underlying.

A call option can be viewed as a mean of acquiring the underlying, whereas a put option can be viewed as a means of selling the underlying. Thus, a call option is logically worth more if the underlying is worth more, and a put option is logically worth more if the underlying is worth less.

The value of the underlying also forms one of the boundaries for calls. The value of a call option cannot exceed the value of the underlying. After all, a call option is only a means of acquiring the underlying. It can never give the holder more benefit than the underlying. Hence, the value of the underlying forms an upper boundary on what a call is worth. The underlying does not provide an upper or lower boundary for puts. That role is played by the exercise price, as we will see in the next section.

To recap what we learned here,

> *The value of a European call option is directly related to the value of the underlying.*
>
> *The value of a European put option is inversely related to the value of the underlying.*

8.1.3 *Effect of the Exercise Price*

The exercise price is a critical factor in determining the value of an option. The exercise price is the hurdle beyond which the underlying must go to justify exercise. For a call, the underlying must rise above the exercise price, and for a put, the underlying must fall below the exercise price, to justify exercise. When the underlying is beyond the exercise price in the appropriate direction (higher for a call, lower for a put), the option is said to be **in the money**. When the underlying is precisely at the exercise price, the option is said to be **at the money**. When the underlying has not reached the exercise price (currently lower for a call, higher for a put), the option is said to be **out of the money**. This characterization of whether the option is in-, at-, or out-of-the-money is referred to as the option's **moneyness**.

For a call option, a lower exercise price has two benefits. One is that there are more values of the underlying at expiration that are above the exercise price, meaning that there are more outcomes in which the call expires in-the-money. The other benefit is that assuming the call expires in-the-money, for any value of the underlying, the option value is greater the lower the exercise price. In other words, at expiration the underlying value S_T will be above the exercise price far more often, the lower is X. And if S_T is indeed higher than X, the payoff of $S_T - X$ is greater, the lower is X.

For puts, the effect is just the opposite. To expire in-the-money, the value of the underlying must fall below the exercise price. The higher the exercise price, the better chance the underlying has of getting below it. Likewise, if the value of the underlying does fall below the exercise price, the higher the exercise price, the greater the payoff. So, if X is higher, S_T will be below it more often, and if S_T is less than X, the payoff of $X - S_T$ is greater, the higher is X for whatever value of S_T occurs.

The exercise price also helps form an upper bound for the value of a European put. If you were holding a European put, the best outcome you could hope for is a zero value of the underlying. For equities, that would mean complete failure and dissolution of the company with shareholders receiving no final payment.[15] In that case, the put would pay $X - S_T$, but with S_T at zero, the put would pay X. If the underlying value goes to zero during the life of the European put, however, the holder cannot collect that payoff until expiration. Nonetheless, the holder would have a risk-free claim on a payoff of X at expiration. Thus, the most the put would be worth is the present value of X, meaning X discounted from expiration to the current day at the risk-free rate.[16] Although the holder cannot collect the payoff by exercising the option, he could sell it for the present value of X.

To recap these results,

The value of a European call option is inversely related to the exercise price.

The value of a European put option is directly related to the exercise price.

8.1.4 *Effect of Time to Expiration*

Logic suggests that longer-term options should be worth more than shorter-term options. That statement is usually true but not always. A call option unquestionably benefits from additional time. For example, the right to buy an asset for $50 is worth a lot more if that right is available for two years instead of one. The additional time provides further opportunity for the underlying to rise above the exercise price. Although that means there is also additional time for the underlying to fall below the exercise price, it hardly matters to the holder of the call because the loss on the downside is limited to the premium paid.

For a European put option, the additional time still provides more opportunity for the underlying price to fall below the exercise price, but with the additional risk of it rising above the exercise price mitigated by the limited loss of the premium if the put expires out-of-the-money. Thus, it sounds as if puts benefit from longer time, but that is not necessarily true. There is a subtle penalty for this additional time. Put option holders are awaiting the sale of the underlying, for which they will receive the exercise price. The longer they have to wait, the lower the present value of the payoff. For some puts, this negative effect can dwarf the positive effect. This situation occurs with a put the longer the time to expiration, the higher the risk-free rate of interest, and the deeper it is in-the-money. The positive effect of time, however, is somewhat more dominant.

Note that we did not mention this effect for calls. For calls, the holder is waiting to pay out money at expiration. More time lowers the value of this possible outlay. Hence, a longer time period helps call option buyers in this regard.

To recap these results,

The value of a European call option is directly related to the time to expiration.

---

15 You might think this point means that people who buy puts are hoping the company goes bankrupt, a seemingly morbid motivation. Yet, put buyers are often people who own the stock and buy the put for protection. This motivation is no different from owning a house and buying fire insurance. You do not want the house to burn down. If your sole motivation in buying the insurance were to make a profit on the insurance, you would want the house to burn down. This moral hazard problem illustrates why it is difficult, if not impossible, to buy insurance on a house you do not own. Likewise, executives are prohibited from owning puts on their companies' stock. Individual investors can own puts on stocks they do not own, because they cannot drive the stock price down.

16 For the put holder to truly have a risk-free claim on X at expiration, given zero value of the underlying today, the underlying value must go to zero and have no possibility of any recovery. If there is any possibility of recovery, the underlying value would not go to zero, as is often observed when a legal filing for bankruptcy is undertaken. Many equities do recover. If there were some chance of recovery but the equity value was zero, demand for the stock would be infinite, which would push the price up.

> *The value of a European put option can be either directly or inversely related to the time to expiration. The direct effect is more common, but the inverse effect can prevail with a put the longer the time to expiration, the higher the risk-free rate, and the deeper it is in-the-money.*

8.1.5 *Effect of the Risk-Free Rate of Interest*

We have already alluded to the effect of the risk-free rate. For call options, a longer time to expiration means that the present value of the outlay of the exercise price is lower. In other words, with a longer time to expiration, the call option holder continues to earn interest on the money that could be expended later in paying the exercise price upon exercise of the option. If the option is ultimately never exercised, this factor is irrelevant, but it remains at best a benefit and at worst has no effect. For puts, the opposite argument prevails. A longer time to expiration with a higher interest rate lowers the present value of the receipt of the exercise price upon exercise. Thus, the value today of what the put holder might receive at expiration is lower. If the put is ultimately never exercised, the risk-free rate has no effect. Thus, at best, a higher risk-free rate has no effect on the value of a put. At worst, it decreases the value of the put.

These results are summarized as follows:

> *The value of a European call is directly related to the risk-free interest rate.*
>
> *The value of a European put is inversely related to the risk-free interest rate.*

8.1.6 *Effect of Volatility of the Underlying*

In studying the pricing of equities, we are conditioned to believe that volatility has a negative effect. After all, investors like return and dislike risk. Volatility is certainly an element of risk. Therefore, volatility is bad for investors, right? Well, partially right.

First, not all volatility is bad for investors. Unsystematic volatility should be irrelevant. Investors can hold diversified portfolios. Systematic volatility is clearly undesirable, but do not think that this means that volatility should be completely avoided where possible. If volatility were universally undesirable, no one would take risks. Clearly risks have to be taken to provide opportunity for reward.

With options, volatility of the underlying is, however, universally desirable. The greater the volatility of the underlying, the more an option is worth. This seemingly counterintuitive result is easy to understand with a little explanation.

First, let us make sure we know what volatility really means. In studying asset returns, we typically represent volatility with the standard deviation of the return, which measures the variation from the average return. The S&P 500 Index has an approximate long-run volatility of around 20%. Under the assumption of a normal distribution, a standard deviation of 20% implies that about 68% of the time, the returns will be within plus or minus one standard deviation of the average. About 95% of the time, they will be within plus or minus two standard deviations of the average. About 99% of the time, they will be within plus or minus three standard deviations of the average. When the distribution is non-normal, different interpretations apply, and in some extreme cases, the standard deviation can be nearly impossible to interpret.

Standard deviation is not the only notion of volatility, however, and it is not even needed at this point. You can proceed fairly safely with a measure as simple as the highest possible value minus the lowest, known as the range. The only requirement we need right now is that the concept of volatility reflects dispersion—how high and how low the underlying can go.

So, regardless of how we measure volatility, the following conditions will hold:

1 A call option will have a higher payoff the higher the underlying is at expiration.

2 A call option will have a zero payoff if it expires with the underlying below the exercise price.

If we could impose greater volatility on the underlying, we should be able to see that in Condition 1, the payoff has a better chance of being greater because the underlying has a greater possibility of large positive returns. In Condition 2, however, the zero payoff is unaffected if we impose greater volatility. Expiring more out-of-the-money is not worse than expiring less out-of-the-money, but expiring more in-the-money is better than expiring less-in-the-money.[17]

For puts, we have

1 A put option will have a higher payoff the lower the underlying is at expiration.

2 A put option will have a zero payoff if it expires with the underlying above the exercise price.

If we could impose greater volatility, we would find that it would have a beneficial effect in (1) because a larger positive payoff would have a greater chance of occurring. In (2), the zero payoff is unaffected. The greater of the option expiring more out-of-the-money is irrelevant. Expiring more out-of-the-money is not worse than expiring less out-of-the-money.

Thus, we summarize our results in this section as

> *The value of a European call is directly related to the volatility of the underlying.*
>
> *The value of a European put is directly related to the volatility of the underlying.*

The combined effects of time and volatility give rise to the concept of the time value of an option. The **time value** of an option is the difference between the market price of the option and its intrinsic value. It represents the market valuation of the potential for higher exercise value relative to the potential for lower exercise value given the volatility of the underlying. Time value of an option is not to be confused with the time value of money, which is the notion of money later being worth less than money today as a result of the combined effects of time and interest. Time value results in an option price being greater with volatility and time but declining as expiration approaches. At expiration, no time value remains and the option is worth only its exercise value. As such, an option price is said to decay over time, a process characterized as **time value decay**, which is covered in more advanced material.

8.1.7 *Effect of Payments on the Underlying and the Cost of Carry*

We previously discussed how payments on the underlying and carrying costs enter into the determination of forward prices. They also affect option prices. Payments on the underlying refer to dividends on stocks and interest on bonds. In addition, some commodities offer a convenience yield benefit. Carrying costs include the actual physical costs of maintaining and/or storing an asset.

Let us first consider the effect of benefits. Payments of dividends and interest reduce the value of the underlying. Stocks and bonds fall in value as dividends and interest are paid. These benefits to holders of these securities do not flow to holders of options. For call option holders, this reduction is a negative factor. The price of

---

17 Think of an option expiring out-of-the-money as like it being dead. (Indeed, the option is dead.) Being "more dead" is not worse than being "less dead."

the underlying is hurt by such payments, and call holders do not get to collect these payments. For put holders, the effect is the opposite. When the value of the underlying is reduced, put holders are helped.

Carrying costs have the opposite effect. They raise the effective cost of holding or shorting the asset. Holding call options enables an investor to participate in movements of the underlying without incurring these costs. Holding put options makes it more expensive to participate in movements in the underlying than by short selling because short sellers benefit from carrying costs, which are borne by owners of the asset.

To summarize the results from this section,

> *A European call option is worth less the more benefits that are paid by the underlying and worth more the more costs that are incurred in holding the underlying.*
>
> *A European put option is worth more the more benefits that are paid by the underlying and worth less the more costs that are incurred in holding the underlying.*

9 LOWER LIMITS FOR PRICES OF EUROPEAN OPTIONS

k identify the factors that determine the value of an option and explain how each factor affects the value of an option;

What we have learned so far forms a framework for understanding how European options are priced. Let us now go a step further and establish a minimum price for these options.

First, we need to look at a call option as similar to the purchase of the underlying with a portion of the purchase price financed by borrowing. If the underlying is a stock, this transaction is usually called a margin transaction. Assume that the underlying is worth S_0. Also assume that you borrow cash in the amount of the present value of X, promising to pay X back T periods later at an interest rate of r. Thus, $X/(1 + r)^T$ is the amount borrowed, and X is the amount to be paid back. Now move forward to time T and observe the price of the underlying, S_T. Upon paying back the loan, the overall strategy will be worth $S_T – X$, which can be positive or negative.

Next, consider an alternative strategy of buying a call option expiring at T with an exercise price of X, the same value as the face value of the loan. We know that the option payoffs will be $S_T – X$ if it expires in-the- money ($S_T > X$) and zero if not ($S_T \le X$). Exhibit 11 compares these two strategies.[18]

18 Note in Exhibit 11, and in others to come, that the inequality ≤ is referred to as out-of-the-money. The case of equality is technically referred to as at-the-money but the verbiage is simplified if we continue to call it out-of-the-money. It is certainly not in-the-money and at-the-money is arguably the same as out-of-the-money. Regardless of one's preference, the equality case can be attached to either of the two outcomes with no effect on our conclusions.

Exhibit 11 Call Option vs. Leveraged (Margin) Transaction

	Outcome at *T*	
	Call Expires Out-of-the-Money ($S_T \le X$)	**Call Expires In-the-Money ($S_T > X$)**
Call	0	$S_T - X$
Leveraged transaction		
Asset	S_T	S_T
Loan	$-X$	$-X$
Total	$S_T - X$	$S_T - X$

When the call expires in-the-money, both transactions produce identical payoffs. When the call expires out-of-the-money, the call value is zero, but the leveraged transaction is almost surely a loss. Its value $S_T - X$ is negative or zero at best (if S_T is exactly equal to X).

If two strategies are found to produce equivalent results in some outcomes but one produces a better result in all other outcomes, then one strategy dominates the former. Here we see that the call strategy dominates the leveraged strategy. Any strategy that dominates the other can never have a lower value at any time. Why would anyone pay more for one strategy than for another if the former will never produce a better result than the latter? Thus, the value of the call strategy, c_0, has to be worth at least the value of the leveraged transaction, S_0 (the value of the asset), minus $X/(1 + r)^T$ (the value of the loan). Hence, $c_0 \ge S_0 - X/(1 + r)^T$.

The inequality means that this statement provides the lowest price of the call, but there is one more thing we need to do. It can easily be true that $X/(1 + r)^T > S_0$. In that case, we are saying that the lowest value is a negative number, but that statement is meaningless. A call can never be worth less than zero, because its holder cannot be forced to exercise it. Thus, we tend to express this relationship as

$$c_0 \ge \text{Max}\left[0, S_0 - X/(1 + r)^T\right] \quad \textbf{(10)}$$

which represents the greater of the value of zero or the underlying price minus the present value of the exercise price. This value becomes the lower limit of the call price.

Now consider an analogous result for puts. Suppose we want to profit from a declining price of the underlying. One way to do this is to sell the underlying short. Suppose we do that and invest cash equal to the present value of X into risk-free bonds that pay X at time T. At time T, given a price of the underlying of S_T, the short sale pays off $-S_T$, a reflection of the payment of S_T to cover the short sale. The bonds pay X. Hence, the total payoff is $X - S_T$.

Now, compare that result with the purchase of a put expiring at T with exercise price of X. If the put expires in–the-money ($S_T < X$), it is worth $X - S_T$. If it expires out-of-the-money ($S_T \ge X$), it is worth zero. Exhibit 12 illustrates the comparison of the put with the short sale and bond strategy. We see that for the in-the-money case, the put and short sale and bond strategies match each other. For the out-of-the-money case, however, the put performs better because the short sale and bond strategy pays $X - S_T$. With $S_T \ge X$, this payment amount is negative. With the put dominating the short sale and bond strategy, the put value cannot be less than the value of the short sale and bond strategy, meaning $p_0 \ge X/(1 + r)^T - S_0$. But as with calls, the right-hand side can be negative, and it hardly helps us to say that a put must sell for more than a negative number. A put can never be worth less than zero, because its owner cannot be forced to exercise it. Thus, the overall result is expressed succinctly as

$$p_0 \geq \text{Max}\left[0, X/(1+r)^T - S_0\right] \tag{11}$$

Exhibit 12 Put vs. Short Sale and Bond Purchase

	Outcome at *T*	
	Put Expires In-the-Money ($S_T < X$)	**Put Expires Out-of-the-Money ($S_T \geq X$)**
Put	$X - S_T$	0
Short sale and bond purchase		
Short sale	$-S_T$	$-S_T$
Bond	X	X
Total	$X - S_T$	$X - S_T$

Let us look at some basic examples. Assume the exercise price is €60, the risk-free rate is 4%, and the expiration is nine months, so T = 9/12 = 0.75. Consider two cases:

Underlying: S_0 = €70

Minimum call price = Max[0,€70 – €60/(1.04)$^{0.75}$] = Max(0,€11.74) = €11.74

Minimum put price = Max[0,€60/(1.04)$^{0.75}$ – €70] = Max(0,–€11.74) = €0.00

Underlying: S_0 = €50

Minimum call price = Max[0,€50 – €60/(1.04)$^{0.75}$] = Max(0,–€8.26) = €0.00

Minimum put price = Max[0,€60/(1.04)$^{0.75}$ – €50] = Max(0,€8.26) = €8.26

To recap, in this section we have established lower limits for call and put option values. Formally restating these results in words,

The lowest value of a European call is the greater of zero or the value of the underlying minus the present value of the exercise price.

The lowest value of a European put is the greater of zero or the present value of the exercise price minus the value of the underlying.

EXAMPLE 6

Basic Principles of European Option Pricing

1 Which of the following factors does *not* affect the value of a European option?

A The volatility of the underlying

B Dividends or interest paid by the underlying

C The percentage of the investor's assets invested in the option

2 Which of the following statements imply that a European call on a stock is worth more?

   A Less time to expiration

   B A higher stock price relative to the exercise price

   C Larger dividends paid by the stock during the life of the option

3 Why might a European put be worth less the longer the time to expiration?

   A The cost of waiting to receive the exercise price is higher.

   B The risk of the underlying is lower over a longer period of time.

   C The longer time to expiration means that the put is more likely to expire out-of-the-money.

4 The loss in value of an option as it moves closer to expiration is called what?

   A Time value decay

   B Volatility diminution

   C Time value of money

5 How does the minimum value of a call or put option differ from its exercise value?

   A The exercise price is adjusted for the time value of money.

   B The minimum value reflects the volatility of the underlying.

   C The underlying price is adjusted for the time value of money.

Solution to 1:

C is correct. The investor's exposure to the option is not relevant to the price one should pay to buy or ask to sell the option. Volatility and dividends or interest paid by the underlying are highly relevant to the value of the option.

Solution to 2:

B is correct. The higher the stock price and the lower the exercise price, the more valuable is the call. Less time to expiration and larger dividends reduce the value of the call.

Solution to 3:

A is correct. Although the longer time benefits the holder of the option, it also has a cost in that exercise of a longer-term put comes much later. Therefore, the receipt of the exercise price is delayed. Longer time to expiration does not lower the risk of the underlying. The longer time also does not increase the likelihood of the option expiring out-of-the-money.

Solution to 4:

A is correct. An option has time value that decays as the expiration approaches. There is no such concept as volatility diminution. Time value of money relates only to the value of money at one point in time versus another.

Solution to 5:

A is correct. The minimum value formula is the greater of zero or the difference between the underlying price and the present value of the exercise price, whereas the exercise value is the maximum of zero and the appropriate difference between the underlying price and the exercise price. Volatility does not affect the minimum price. It does not make sense to adjust the underlying price for the time value of money for the simple reason that it is already adjusted for the time value of money.

10 PUT-CALL PARITY, PUT-CALL-FORWARD PARITY

l. explain put–call parity for European options;

m. explain put–call–forward parity for European options;

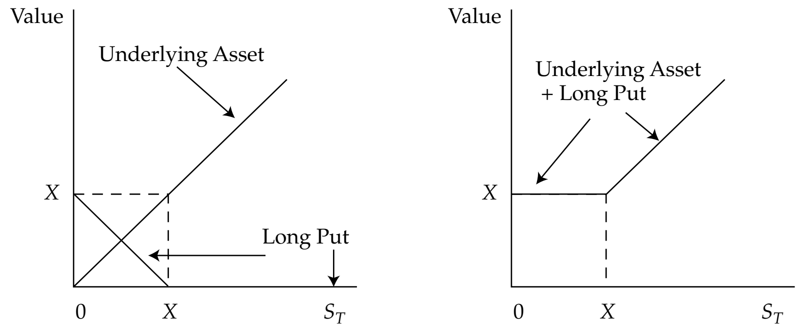
One of the first concepts that a trader learns in options is the parity relationship between puts and calls. Even though the word "parity" means "equivalence," puts and calls are not equivalent. There is, however, a relationship between the call price and the price of its corresponding put, which we refer to as put–call parity.

Suppose Investor A owns an asset that has a current price of S_0. Assume the asset makes no cash payments and has no carrying costs. The end of the holding period is time T, at which point the asset will be worth S_T. Fearing the possibility that S_T will decline, Investor A buys a put option with an exercise price of X, which can be used to sell the asset for X at time T. This put option has a premium of p_0. Combined with the value of the asset, the investor's current position is worth $S_0 + p_0$, which is the investor's money at risk. This strategy of holding the asset and a put is sometimes called a **protective put**.

At expiration, the value of the asset is S_T. The value of the put will be either zero or $X - S_T$. If the asset increases in value such that $S_T \geq X$, then the overall position is worth S_T. The asset has performed well, and the investor will let the put expire. If the asset value declines to the point at which $S_T < X$, the asset is worth S_T, and the put is worth $X - S_T$, for a total of X. In other words, the investor would exercise the put, selling the asset for X, which exceeds the asset's current value of S_T.

This strategy seems like a reasonable and possibly quite attractive investment. Investor A receives the benefit of unlimited upside potential, with the downside performance truncated at X. Exhibit 13 shows the performance of the protective put. The graph on the left illustrates the underlying asset and the put. The graph on the right shows their combined effects.

Exhibit 13 Protective Put (Asset Plus Long Put)

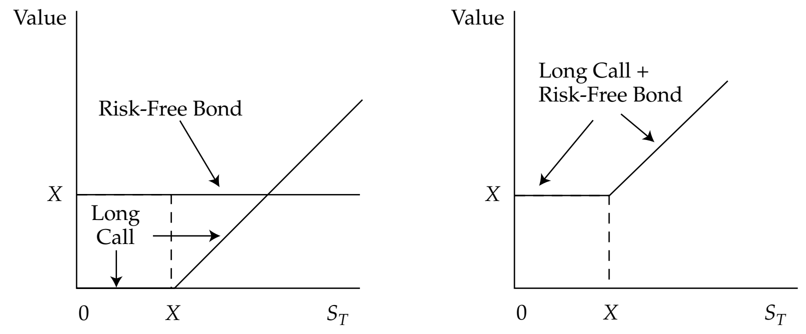


Consider Investor B, an options trader. At time 0, this investor buys a call option on this asset with an exercise price of X that expires at T and a risk-free zero-coupon bond with a face value of X that matures at T. The call costs c_0, and the bond costs the present value of X, which is $X/(1 + r)^T$. Thus, Investor B has invested funds of $c_0 + X/$

$(1 + r)^T$. This strategy is sometimes known as a **fiduciary call**. If the underlying price exceeds the exercise price at expiration, the call will be worth $S_T - X$, and the bond will mature and pay a value of X. These values combine to equal S_T. If the underlying price does not exceed the exercise price at expiration, the call expires worthless and the bond is worth X for a combined value of X.

Exhibit 14 shows the performance of the fiduciary call. The graph on the left shows the call and bond, and the graph on the right shows the combined effects of the two strategies.

Exhibit 14 Fiduciary Call (Long Call Plus Risk-Free Bond)



Comparing Exhibit 13 with Exhibit 14 shows that a protective put and a fiduciary call produce the same result. Exhibit 15 shows this result more directly by identifying the payoffs in the various outcomes. Recall that Investor A committed funds of $S_0 + p_0$, while Investor B committed funds of $c_0 + X/(1 + r)^T$. If both investors receive the same payoffs at time T regardless of the asset price at T, the amounts they invest at time 0 have to be the same. Thus, we require

$$S_0 + p_0 = c_0 + X\big/(1 + r)^T \qquad (12)$$

This relationship is known as **put–call parity**.

Exhibit 15 Protective Put vs. Fiduciary Call

	Outcome at T	
	Put Expires In-the-Money ($S_T < X$)	**Call Expires In-the-Money ($S_T \geq X$)**
Protective put		
Asset	S_T	S_T
Long put	$X - S_T$	0
Total	X	S_T
Fiduciary call		
Long call	0	$S_T - X$
Risk-free bond	X	X
Total	X	S_T

For a simple example, assume call and put options with an exercise price of ¥100,000 in which the underlying is at ¥90,000 at time 0. The risk-free rate is 2% and the options expire in two months, so $T = 2/12 = 0.167$. To completely fill in the puzzle, we would need to know the put or call price, from which we could obtain the other. For now, let us write this relationship as

$$p_0 - c_0 = X/(1 + r)^T - S_0$$

The right side would be ¥100,000/(1.02)$^{0.167}$ – ¥90,000 = ¥9,670. Thus, the put price should exceed the call price by ¥9,670. Thus, if the call were priced at ¥5,000, the put price would be ¥14,670. If we knew the put price, we could obtain the call price. Put–call parity does not tell us which price is correct, and it requires knowledge of one price to get the other. Alternatively, it can tell us the difference in the put and call prices.

Put–call parity must hold, at least within transaction costs, or arbitrage opportunities would arise. For example, suppose Investor C observes market prices and finds that the left- hand side of put–call parity, $S_0 + p_0$, is less than the right-hand side, $c_0 + X/(1 + r)^T$. Thus, the put and the stock cost less than the call and the bond. Knowing that there should be equality (parity), Investor C executes an arbitrage transaction, selling the overpriced transactions (the call and the bond) and buying the underpriced transactions (the asset and the put).[19] By selling the higher priced side and buying the lower priced side, Investor C will take in more money than she will pay out, a net inflow of $c_0 + X/(1 + r)^T - (S_0 + p_0)$. At expiration, the long put and long asset will offset the short call and bond, as shown in Exhibit 16.

Exhibit 16 Put–Call Parity Arbitrage

		Outcome at *T*	
Transaction	**Cash Flow at Time 0**	**Put Expires In-the-Money ($S_T < X$)**	**Call Expires In-the-Money ($S_T \geq X$)**
Buy asset	$-S_0$	S_T	S_T
Buy put	$-p_0$	$X - S_T$	0
Sell call	$+c_0$	0	$-(S_T - X)$
Borrow	$+X/(1 + r)^T$	$-X$	$-X$
Total	$-S_0 - p_0 + c_0 + X/(1 + r)^T > 0$	0	0

In simple terms, if $S_T < X$, the short call expires out-of-the-money and the put is exercised to sell the asset for X. This cash, X, is then used to pay off the loan. The net effect is that no money flows in or out at T. If $S_T \geq X$, the put expires out-of-the money, and the short call is exercised, meaning that Investor C must sell the asset for X. This cash, X, is then used to pay off the loan. Again, no money flows in or out. The net effect is a perfect hedge in which no money is paid out or received at T. But there was money taken in at time 0. Taking in money today and never having to pay it out is an arbitrage profit. Arbitrage opportunities like this, however, will be noticed by many investors who will engage in the same transactions. Prices will adjust until parity is restored, whereby $S_0 + p_0 = c_0 + X/(1 + r)^T$.

19 Selling the bond is equivalent to borrowing, meaning to issue a loan.

Put–call parity provides tremendous insights into option pricing. Recall that we proved that going long the asset and long a put is equivalent to going long a call and long a risk-free bond. We can rearrange the put–call parity equation in the following ways:

$$\begin{aligned}
S_0 + p_0 &= c_0 + X/(1+r)^T \\
&\Rightarrow \\
p_0 &= c_0 - S_0 + X/(1+r)^T \\
c_0 &= p_0 + S_0 - X/(1+r)^T \\
S_0 &= c_0 - p_0 + X/(1+r)^T \\
X/(1+r)^T &= S_0 + p_0 - c_0
\end{aligned}$$

By using the symbols and the signs in these versions of put–call parity, we can see several important interpretations. In the equations below, plus signs mean long and minus signs mean short:

$$\begin{aligned}
p_0 &= c_0 - S_0 + X/(1+r)^T &&\Rightarrow \text{long put} = \text{long call, short asset, long bond} \\
c_0 &= p_0 + S_0 - X/(1+r)^T &&\Rightarrow \text{long call} = \text{long put, long asset, short bond} \\
S_0 &= c_0 - p_0 + X/(1+r)^T &&\Rightarrow \text{long asset} = \text{long call, short put, long bond} \\
X/(1+r)^T &= S_0 + p_0 - c_0 &&\Rightarrow \text{long bond} = \text{long asset, long put, short call}
\end{aligned}$$

You should be able to convince yourself of any of these points by constructing a table similar to Exhibit 15.[20]

10.1 Put–Call–Forward Parity

Recall that we demonstrated that one could create a risk-free position by going long the asset and selling a forward contract.[21] It follows that one can synthetically create a position in the asset by going long a forward contract and long a risk-free bond. Recall our put–call parity discussion and assume that Investor A creates his protective put in a slightly different manner. Instead of buying the asset, he buys a forward contract and a risk-free bond in which the face value is the forward price. Exhibit 17 shows that this strategy is a synthetic protective put. Because we showed that the fiduciary call is equivalent to the protective put, a fiduciary call has to be equivalent to a protective put with a forward contract. Exhibit 18 demonstrates this point.

Exhibit 17 Protective Put with Forward Contract vs. Protective Put with Asset

	Outcome at T	
	Put Expires In-the-Money ($S_T < X$)	**Put Expires Out-of-the-Money ($S_T \geq X$)**
Protective put with asset		
Asset	S_T	S_T
Long put	$X - S_T$	0
Total	X	S_T

(continued)

20 As a further exercise, you might change the signs of each term in the above and provide the appropriate interpretations.
21 You might wish to review Exhibit 6.

Exhibit 17 (Continued)

	Outcome at T	
	Put Expires In-the-Money ($S_T < X$)	**Put Expires Out-of-the-Money ($S_T \geq X$)**
Protective put with forward contract		
Risk-free bond	$F_0(T)$	$F_0(T)$
Forward contract	$S_T - F_0(T)$	$S_T - F_0(T)$
Long put	$X - S_T$	0
Total	X	S_T

Exhibit 18 Protective Put with Forward Contract vs. Fiduciary Call

	Outcome at T	
	Put Expires In-the-Money ($S_T < X$)	**Call Expires In-the-Money ($S_T \geq X$)**
Protective Put with Forward Contract		
Risk-free bond	$F_0(T)$	$F_0(T)$
Forward contract	$S_T - F_0(T)$	$S_T - F_0(T)$
Long put	$X - S_T$	0
Total	X	S_T
Fiduciary Call		
Call	0	$S_T - X$
Risk-free bond	X	X
Total	X	S_T

It follows that the cost of the fiduciary call must equal the cost of the synthetic protective put, giving us what is referred to as **put–call–forward parity**,

$$F_0(T)\big/(1+r)^T + p_0 = c_0 + X\big/(1+r)^T \tag{13}$$

Returning to our put–call parity example, a forward contract on ¥90,000 expiring in two months with a 2% interest rate would have a price of ¥90,000(1.02)$^{0.167}$ = ¥90,298. Rearranging Equation 13, we have

$$p_0 - c_0 = \left[X - F_0(T)\right]\big/(1+r)^T$$

The right-hand side is (¥100,000 – ¥90,298)/(1.02)$^{0.167}$ = ¥9,670, which is the same answer we obtained using the underlying asset rather than the forward contract. Naturally these two models give us the same answer. They are both based on the assumption that no arbitrage is possible within the spot, forward, and options markets.

So far we have learned only how to price options in relation to other options, such as a call versus a put or a call or a put versus a forward. We need a way to price options versus their underlying.

EXAMPLE 7

Put–Call Parity

1 Which of the following statements *best* describes put–call parity?
  A The put price always equals the call price.
  B The put price equals the call price if the volatility is known.
  C The put price plus the underlying price equals the call price plus the present value of the exercise price.

2 From put–call parity, which of the following transactions is risk-free?
  A Long asset, long put, short call
  B Long call, long put, short asset
  C Long asset, long call, short bond

Solution to 1:

C is correct. The put and underlying make up a protective put, while the call and present value of the exercise price make up a fiduciary call. The put price equals the call price for certain combinations of interest rates, times to expiration, and option moneyness, but these are special cases. Volatility has no effect on put–call parity.

Solution to 2:

A is correct. The combination of a long asset, long put, and short call is risk free because its payoffs produce a known cash flow of the value of the exercise price. The other two combinations do not produce risk-free positions. You should work through the payoffs of these three combinations in the form of Exhibit 12.

BINOMIAL VALUATION OF OPTIONS

11

n explain how the value of an option is determined using a one-period binomial model;

Because the option payoff is determined by the underlying, if we know the outcome of the underlying, we know the payoff of the option. That means that the price of the underlying is the only element of uncertainty. Moreover, the uncertainty is not so much the value of the underlying at expiration as it is whether the underlying is above or below the exercise price. If the underlying is above the exercise price at expiration, the payoff is $S_T - X$ for calls and zero for puts. If the underlying is below the exercise price at expiration, the payoff is zero for calls and $X - S_T$ for puts. In other words, the payoff of the option is straightforward and known, as soon as we know whether the option expires in- or out-of-the-money. Note that for forwards, futures, and swaps, there is no such added complexity. The payoff formula is the same regardless of whether the underlying is above or below the hurdle.

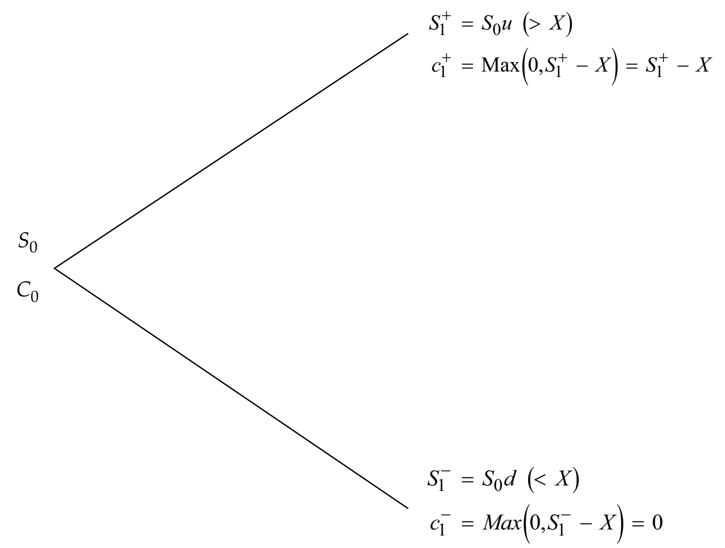
As a result of this characteristic of options, derivation of an option pricing model requires the specification of a model of a random process that describes movements in the underlying. Given the entirely different nature of the payoffs above and below the exercise price, it might seem difficult to derive the option price, even if we could model movements in the underlying. Fortunately, the process is less difficult than it first appears.

At this level of treatment, we will start with a very simple model that allows only two possible movements in the underlying—one going up and one going down from where it is now. This model with two possible outcomes is called the **binomial model**. Start with the underlying at S_0, and let it go up to S_1^+ or down to S_1^-. We cannot arbitrarily set these values at just anything. We will be required to know the values of S_1^+ and S_1^-. That does not mean we know which outcome will occur. It means that we know only what the possibilities are. In doing so, we effectively know the volatility. Assume the probability of the move to S_1^+ is q and the probability of the move to S_1^- is $1 - q$. We specify the returns implied by these moves as up and down factors, u and d, where

$$u = \frac{S_1^+}{S_0},\ d = \frac{S_1^-}{S_0} \tag{14}$$

Now, consider a European call option that expires at time 1 and has an exercise price of X. Let the call prices be c_0 today and c_1^+ and c_1^- at expiration. Exhibit 19 illustrates the model. Our objective is to determine the price of the option today, meaning to determine a formula for c_0. Knowing what we know about arbitrage and the pricing of forward contracts, it would seem we could construct a risk-free portfolio involving this option.

Exhibit 19 The Binomial Option Pricing Model



Because call options and the underlying move together, one possibility is that buying the underlying and selling a call could create a hedge. Indeed it does, but one unit of each is not the appropriate balance. Let us sell one call and hold h units of the underlying. The value h is unknown at the moment, but we will be able to determine its value. The value today of a combination of h units of the underlying and one short call is

$$V_0 = hS_0 - c_0$$

Think of V_0 as the amount of money invested. Depending on which of the two paths is taken by the underlying, the value of this portfolio at time 1 will be

$$V_1^+ = hS_1^+ - c_1^+$$
$$\text{or} \tag{15}$$
$$V_1^- = hS_1^- - c_1^-$$

If the portfolio were hedged, then V_1^+ would equal V_1^-. We can set V_1^+ and V_1^- equal to each other and solve for the value of h that assures us that the portfolio is hedged:

$$V_1^+ = V_1^-$$
$$\Rightarrow hS_1^+ - c_1^+ = hS_1^- - c_1^- \tag{16}$$
$$\Rightarrow h = \frac{c_1^+ - c_1^-}{S_1^+ - S_1^-}$$

The values on the right-hand side are known, so we can easily calculate h. Thus, we can derive the number of units of the underlying that will perfectly hedge one unit of the short call.

We know that a perfectly hedged investment should earn the risk-free rate, r. Thus, the following statement must be true:

$$V_1^+ \text{ (or } V_1^-\text{)} = V_0(1 + r)$$

We can substitute the value of V_1^+ or V_1^- from Equation 15 into the above equation. Then we do a little algebra, which is not important to this discussion, and obtain the formula for the option price,

$$c_0 = \frac{\pi c_1^+ + (1 - \pi)c_1^-}{1 + r}$$
$$\text{where} \tag{17}$$
$$\pi = \frac{1 + r - d}{u - d}$$

Equation 17 shows that the value of the call today is a weighted average of the next two possible call prices at expiration, where the weights, π and $1 - \pi$, are given by the second formula in Equation 17.

This formula sheds a great deal of light on option pricing. Notice the following:

- The volatility of the underlying, which is reflected in the difference between S_1^+ and S_1^- and affects c_1^+ and c_1^-, is an important factor in determining the value of the option.
- The probabilities of the up and down moves, q and $1 - q$, do not appear in the formula.[22]
- The values π and $1 - \pi$ are similar to probabilities and are often called synthetic or pseudo probabilities. They produce a weighted average of the next two possible call values, a type of expected future value.
- The formula takes the form of an expected future value, the numerator, discounted at the risk-free rate.

22 We introduced them earlier to help make this point, but ultimately they serve no purpose.

On the first point, if volatility increases, the difference between S_1^+ and S_1^- increases, which widens the range between c_1^+ and c_1^-, leading to a higher option value. The upper payoff, c_1^+, will be larger and the lower payoff, c_1^-, will still be zero.[23] On the second point, the actual probabilities of the up and down moves do not matter. This result is because of our ability to construct a hedge and the rule of arbitrage. On the third point, the irrelevance of the actual probabilities is replaced by the relevance of a set of synthetic or pseudo probabilities, π and 1 – π, which are called **risk-neutral probabilities**. On the fourth point, these risk-neutral probabilities are used to find a synthetic expected value, which is then discounted at the risk-free rate. Thus, the option is valued as though investors are risk neutral. As we discussed extensively earlier, that is not the same as assuming that investors are risk neutral.

If the option does not trade at the specified formula, Equation 17, investors can engage in arbitrage transactions. If the option is trading too high relative to the formula, investors can sell the call, buy h shares of the underlying, and earn a return in excess of the risk-free rate, while funding the transaction by borrowing at the risk-free rate. The combined actions of arbitrageurs will result in downward pressure on the option price until it converges to the model price. If the option price is too low, buying the call, selling short h units of the asset, and investing the proceeds in risk-free bonds will generate risk-free cash that will earn more than the risk-free rate. The combined actions of arbitrageurs doing this will pressure the call price to rise until it reaches the price given by the model.

We will omit the details, but the hedge portfolio can also be constructed with puts.[24] Changing the c's to p's leads to the binomial put option pricing formula,

$$p_0 = \frac{\pi p_1^+ + (1 - \pi) p_1^-}{1 + r} \tag{18}$$

with the risk-neutral probability π determined by the same formula as for calls, as shown in Equation 17.

Let us construct a simple example. Let S_0 be £40 and the risk-free rate be 5%. The up and down factors are u = 1.20 and d = 0.75. Thus, the next two possible prices of the asset are S_1^+ = £40(1.20) = £48 and S_1^- = £40(0.75) = £30. Consider a call and a put that have exercise prices of £38. Then the next two possible values of the call and put are

$$c_1^+ = \text{Max}(0, £48 - £38) = £10$$

$$c_1^- = \text{Max}(0, £30 - £38) = £0$$

$$p_1^+ = \text{Max}(0, £38 - £48) = £0$$

$$p_1^- = \text{Max}(0, £38 - £30) = £8$$

Next we compute the risk-neutral probability,

$$\pi = \frac{1 + 0.05 - 0.75}{1.20 - 0.75} = 0.667$$

23 Although the lower payoff is zero in this example, that will not always be the case.

24 A long position in h units of the underlying would be hedged with one long put. The formula for h is the same as the one given here for calls, with call prices in the numerator instead of put prices.

The values of the call and put are

$$c_0 = \frac{0.667(£10) + (1 - 0.667)£0}{1.05} = £6.35$$

and

$$p_0 = \frac{0.667(£0) + (1 - 0.667)£8}{1.05} = £2.54$$

The binomial model, as we see it here, is extremely simple. In reality, of course, there are more than two possible next-period prices for the underlying. As it turns out, we can extend the number of periods and subdivide an option's life into an increasing number of smaller time periods. In that case, we can obtain a more accurate and realistic model for option pricing, one that is widely used in practice. Given our objective in this reading of understanding the basic ideas behind derivative pricing, the one-period model is sufficient for the time being.

EXAMPLE 8

Binomial Valuation of Options

1 Which of the following terms directly represents the volatility of the underlying in the binomial model?
   - **A** The standard deviation of the underlying
   - **B** The difference between the up and down factors
   - **C** The ratio of the underlying value to the exercise price.

2 Which of the following is *not* a factor in pricing a call option in the binomial model?
   - **A** The risk-free rate
   - **B** The exercise price
   - **C** The probability that the underlying will go up

3 Which of the following *best* describes the binomial option pricing formula?
   - **A** The expected payoff is discounted at the risk-free rate plus a risk premium.
   - **B** The spot price is compounded at the risk-free rate minus the volatility premium.
   - **C** The expected payoff based on risk-neutral probabilities is discounted at the risk-free rate.

Solution to 1:

B is correct. The up and down factors express how high and how low the underlying can go. Standard deviation does not appear directly in the binomial model, although it is implicit. The ratio of the underlying value to the exercise price expresses the moneyness of the option.

Solution to 2:

C is correct. The actual probabilities of the up and down moves are irrelevant to pricing options. The risk-free and exercise price are, of course, highly relevant.

Solution to 3:

C is correct. Risk-neutral probabilities are used, and discounting is at the risk-free rate. There is no risk premium incorporated into option pricing because of the use of arbitrage.

We have now seen how to obtain the price of a European option. Let us now consider what happens if the options are American, meaning they have the right to be exercised early.

12 AMERICAN OPTION PRICING

- explain under which circumstances the values of European and American options differ.

First, we will use upper case letters for American call and put prices: C_0 and P_0. Second, we know that American options possess every characteristic of European options and one additional trait: They can be exercised at any time prior to expiration. Early exercise cannot be required, so the right to exercise early cannot have negative value. Thus, American options cannot sell for less than European options. Thus, we can state the following:

$$\begin{aligned} C_0 &\geq c_0 \\ P_0 &\geq p_0 \end{aligned} \tag{19}$$

Given the price of the underlying at S_0, the early-exercise feature means that we can exercise the option at any time. So, we can claim the value Max(0,S_0 – X) for calls and Max(0,X – S_0) for puts. These values establish new minimum prices for American calls and puts,

$$\begin{aligned} C_0 &= \text{Max}(0, S_0 - X) \\ P_0 &= \text{Max}(0, X - S_0) \end{aligned} \tag{20}$$

For call options, we previously learned that a European call has a minimum value given by Equation 10, which is restated here:

$$c_0 \geq \text{Max}\left[0, S_0 - X/(1+r)^T\right]$$

Comparing Max(0,S_0 – X) (the minimum for American calls) with Max[0,S_0 – X/(1 + r)^T] (the minimum for European calls) reveals that the latter is either the same or higher. There are some circumstances in which both minima are zero, some in which the American minimum is zero and the European minimum is positive, and some in which both are positive, in which case $S_0 - X/(1 + r)^T$ is unquestionably more than $S_0 - X$. Given that an American call price cannot be less than a European call price, we have to reestablish the American call minimum as Max[0,S_0 – X/(1 + r)^T].

For put options, we previously learned that a European put has a minimum value given by Equation 11, which is restated here:

$$p_0 \geq \text{Max}\left[0, X/(1+r)^T - S_0\right]$$

Comparing Max(0,X – S_0) (the minimum for American puts) with Max[0,X/(1 + r)^T – S_0] (the minimum for European puts) reveals that the former is never less. In some circumstances, they are both zero. In some, $X - S_0$ is positive and X/(1 + r)

$^{T} - S_0$ is negative, and in some cases both are positive but $X - S_0$ is unquestionably more than $X/(1 + r)^T - S_0$. Thus, the American put minimum value is the exercise value, which is Max$(0,X - S_0)$.

So, now we have new minimum prices for American calls and puts:

$$C_0 \geq \text{Max}\left[0, S_0 - X/(1+r)^T\right]$$
$$P_0 \geq \text{Max}(0, X - S_0) \quad (21)$$

Thus, in the market these options will trade for at least these values.

Let us return to the previous examples for the minimum values. The exercise price is €60, the risk-free rate is 4%, and the expiration is T = 0.75. Consider the two cases below:

Underlying: S_0 = €70

- The minimum European call price was previously calculated as €11.74. The exercise value of the American call is Max(0,€70 – €60) = €10. The American call has to sell for at least as much as the European call, so the minimum price of the American call is €11.74.
- The minimum European put price was €0.00. This is also the exercise value of the American put [Max(0,€60 – €70) = €0.00], so the minimum price of the American put is still €0.00

Underlying: S_0 = €50

- The minimum European call price was previously calculated as €0.00. The exercise value of the American call is Max(0,€50 – €60) = €0.00, so €0.00 is still the minimum price of the American call.
- The minimum European put price was previously calculated as €8.26. The exercise value of the American put is Max(0,€60 – €50) = €10. So, €10 is the minimum price of the American put.

The call result leads us to a somewhat surprising conclusion. With the exception of what happens at expiration when American and European calls are effectively the same and both worth the exercise value, an American call is always worth more in the market than exercised. That means that an American call will never be exercised early. This result is probably not intuitive.

Consider a deep in-the-money call. One might think that if the holder expected the underlying to not increase any further, exercise might be justified. Yet, we said the call would sell for more in the market than its exercise value. What is the rationale? If the investor thinks the underlying will not go up any further and thus expects no further gains from the option, why would she prefer the underlying? Would the investor be happier holding the underlying, which she believes is not expected to increase? Moreover, she would tie up more funds exercising to acquire the underlying than if she just held on to the option or, better yet, sold it to another investor.

So far, however, we have left out a possible factor that can affect early exercise. Suppose the underlying is a stock and pays dividends. When a stock goes ex-dividend, its price instantaneously falls. Although we will omit the details, an investor holding a call option may find it worthwhile to exercise the call just before the stock goes ex-dividend. The capture of the dividend, thereby avoiding the ex-dividend drop in the price of the underlying, can make early exercise worthwhile. If the underlying is a bond, coupon interest can also motivate early exercise. But if there are significant carrying costs, the motivation for early exercise is weakened. Storage costs lend a preference for owning the option over owning the underlying.

Because the minimum value of an American put exceeds the minimum value of the European put, there is a much stronger motivation for early exercise. Suppose you owned an American put on a stock that is completely bankrupt, with a zero stock price and no possibility of recovery. You can either wait until expiration and capture its exercise value of Max(0,X – S_T) = Max(0,X – 0) = Max(0,X) = X, or you can capture that value by exercising now. Obviously now is better. As it turns out, however, the underlying does not need to go all the way to zero. There is a critical point at which a put is so deep in-the-money that early exercise is justified. This rationale works differently for a call. A deep in-the-money put has a limit to its ultimate value. It can get no deeper than when the underlying goes to zero. For a call, there is no limit to its moneyness because the underlying has no upper limit to its price.

Although dividends and coupon interest encourage early exercise for calls, they discourage early exercise for puts. The loss from the decline in the price of the underlying that is avoided by exercising a call just before the decline works to the benefit of a put holder. Therefore, if a put holder were considering exercising early, he would be better off waiting until right after the dividend or interest were paid. Carrying costs on the underlying, which discourage exercise for calls, encourage exercise for puts.

At this point, we cannot determine the critical prices at which American options are best exercised early. We require more knowledge and experience with option pricing models, which is covered in more advanced material.

EXAMPLE 9

American Option Pricing

1 With respect to American calls, which of the following statements is *most* accurate?

   A American calls should be exercised early if the underlying has reached its expected maximum price.

   B American calls should be exercised early if the underlying has a lower expected return than the risk-free rate.

   C American calls should be exercised early only if there is a dividend or other cash payment on the underlying.

2 The effect of dividends on a stock on early exercise of a put is to:

   A make early exercise less likely.

   B have no effect on early exercise.

   C make early exercise more likely.

Solution to 1:

C is correct. Cash payments on the underlying are the only reason to exercise American calls early. Interest rates, the expected return on the underlying, and any notion of a maximum price is irrelevant. But note that a dividend does not mean that early exercise should automatically be conducted. A dividend is only a necessary condition to justify early exercise for calls.

Solution to 2:

A is correct. Dividends drive down the stock price when the dividend is paid. Thus, all else being equal, a stock paying dividends has a built-in force that drives down the stock price. This characteristic discourages early exercise, because stock price declines are beneficial to holders of puts.

SUMMARY

This reading on derivative pricing provides a foundation for understanding how derivatives are valued and traded. Key points include the following:

- The price of the underlying asset is equal to the expected future price discounted at the risk-free rate, plus a risk premium, plus the present value of any benefits, minus the present value of any costs associated with holding the asset.
- An arbitrage opportunity occurs when two identical assets or combinations of assets sell at different prices, leading to the possibility of buying the cheaper asset and selling the more expensive asset to produce a risk-free return without investing any capital.
- In well-functioning markets, arbitrage opportunities are quickly exploited, and the resulting increased buying of underpriced assets and increased selling of overpriced assets returns prices to equivalence.
- Derivatives are priced by creating a risk-free combination of the underlying and a derivative, leading to a unique derivative price that eliminates any possibility of arbitrage.
- Derivative pricing through arbitrage precludes any need for determining risk premiums or the risk aversion of the party trading the option and is referred to as risk-neutral pricing.
- The value of a forward contract at expiration is the value of the asset minus the forward price.
- The value of a forward contract prior to expiration is the value of the asset minus the present value of the forward price.
- The forward price, established when the contract is initiated, is the price agreed to by the two parties that produces a zero value at the start.
- Costs incurred and benefits received by holding the underlying affect the forward price by raising and lowering it, respectively.
- Futures prices can differ from forward prices because of the effect of interest rates on the interim cash flows from the daily settlement.
- Swaps can be priced as an implicit series of off-market forward contracts, whereby each contract is priced the same, resulting in some contracts being positively valued and some negatively valued but with their combined value equaling zero.
- At expiration, a European call or put is worth its exercise value, which for calls is the greater of zero or the underlying price minus the exercise price and for puts is the greater of zero and the exercise price minus the underlying price.
- European calls and puts are affected by the value of the underlying, the exercise price, the risk-free rate, the time to expiration, the volatility of the underlying, and any costs incurred or benefits received while holding the underlying.
- Option values experience time value decay, which is the loss in value due to the passage of time and the approach of expiration, plus the moneyness and the volatility.
- The minimum value of a European call is the maximum of zero and the underlying price minus the present value of the exercise price.
- The minimum value of a European put is the maximum of zero and the present value of the exercise price minus the price of the underlying.

- European put and call prices are related through put–call parity, which specifies that the put price plus the price of the underlying equals the call price plus the present value of the exercise price.
- European put and call prices are related through put–call–forward parity, which shows that the put price plus the value of a risk-free bond with face value equal to the forward price equals the call price plus the value of a risk-free bond with face value equal to the exercise price.
- The values of European options can be obtained using the binomial model, which specifies two possible prices of the asset one period later and enables the construction of a risk-free hedge consisting of the option and the underlying.
- American call prices can differ from European call prices only if there are cash flows on the underlying, such as dividends or interest; these cash flows are the only reason for early exercise of a call.
- American put prices can differ from European put prices, because the right to exercise early always has value for a put, which is because of a lower limit on the value of the underlying.

PRACTICE PROBLEMS

1 For a risk-averse investor, the price of a risky asset, assuming no additional costs and benefits of holding the asset, is:
  A unrelated to the risk-free rate.
  B directly related to its level of risk.
  C inversely related to its level of risk.

2 An arbitrage opportunity is *least likely* to be exploited when:
  A one position is illiquid.
  B the price differential between assets is large.
  C the investor can execute a transaction in large volumes.

3 An arbitrageur will *most likely* execute a trade when:
  A transaction costs are low.
  B costs of short-selling are high.
  C prices are consistent with the law of one price.

4 An arbitrage transaction generates a net inflow of funds:
  A throughout the holding period.
  B at the end of the holding period.
  C at the start of the holding period.

5 Which of the following combinations replicates a long derivative position?
  A A short derivative and a long asset
  B A long asset and a short risk-free bond
  C A short derivative and a short risk-free bond

6 Most derivatives are priced by:
  A assuming that the market offers arbitrage opportunities.
  B discounting the expected payoff of the derivative at the risk-free rate.
  C applying a risk premium to the expected payoff of the derivative and its risk.

7 The price of a forward contract:
  A is the amount paid at initiation.
  B is the amount paid at expiration.
  C fluctuates over the term of the contract.

8 Assume an asset pays no dividends or interest, and also assume that the asset does not yield any non-financial benefits or incur any carrying cost. At initiation, the price of a forward contract on that asset is:
  A lower than the value of the contract.
  B equal to the value of the contract.
  C greater than the value of the contract.

9 With respect to a forward contract, as market conditions change:
  A only the price fluctuates.
  B only the value fluctuates.
  C both the price and the value fluctuate.

10 The value of a forward contract at expiration is:

A positive to the long party if the spot price is higher than the forward price.

B negative to the short party if the forward price is higher than the spot price.

C positive to the short party if the spot price is higher than the forward price.

11 At the initiation of a forward contract on an asset that neither receives benefits nor incurs carrying costs during the term of the contract, the forward price is equal to the:

A spot price.

B future value of the spot price.

C present value of the spot price.

12 Stocks BWQ and ZER are each currently priced at $100 per share. Over the next year, stock BWQ is expected to generate significant benefits whereas stock ZER is not expected to generate any benefits. There are no carrying costs associated with holding either stock over the next year. Compared with ZER, the one-year forward price of BWQ is *most likely*:

A lower.

B the same.

C higher.

13 If the net cost of carry of an asset is positive, then the price of a forward contract on that asset is *most likely*:

A lower than if the net cost of carry was zero.

B the same as if the net cost of carry was zero.

C higher than if the net cost of carry was zero.

14 If the present value of storage costs exceeds the present value of its convenience yield, then the commodity's forward price is *most likely*:

A less than the spot price compounded at the risk-free rate.

B the same as the spot price compounded at the risk-free rate.

C higher than the spot price compounded at the risk-free rate.

15 Which of the following factors *most likely* explains why the spot price of a commodity in short supply can be greater than its forward price?

A Opportunity cost

B Lack of dividends

C Convenience yield

16 When interest rates are constant, futures prices are *most likely*:

A less than forward prices.

B equal to forward prices.

C greater than forward prices.

17 In contrast to a forward contract, a futures contract:

A trades over-the-counter.

B is initiated at a zero value.

C is marked-to-market daily.

18 To the holder of a long position, it is more desirable to own a forward contract than a futures contract when interest rates and futures prices are:

A negatively correlated.

B uncorrelated.

C positively correlated.

19 The value of a swap typically:

A is non-zero at initiation.

B is obtained through replication.

C does not fluctuate over the life of the contract.

20 The price of a swap typically:

A is zero at initiation.

B fluctuates over the life of the contract.

C is obtained through a process of replication.

21 The value of a swap is equal to the present value of the:

A fixed payments from the swap.

B net cash flow payments from the swap.

C underlying at the end of the contract.

22 If no cash is initially exchanged, a swap is comparable to a series of forward contracts when:

A the swap payments are variable.

B the combined value of all the forward contracts is zero.

C all the forward contracts have the same agreed-on price.

23 For a swap in which a series of fixed payments is exchanged for a series of floating payments, the parties to the transaction:

A designate the value of the underlying at contract initiation.

B value the underlying solely on the basis of its market value at the end of the swap.

C value the underlying sequentially at the time of each payment to determine the floating payment.

24 A European call option and a European put option are written on the same underlying, and both options have the same expiration date and exercise price. At expiration, it is possible that both options will have:

A negative values.

B the same value.

C positive values.

25 At expiration, a European put option will be valuable if the exercise price is:

A less than the underlying price.

B equal to the underlying price.

C greater than the underlying price.

26 The value of a European call option at expiration is the greater of zero or the:

A value of the underlying.

B value of the underlying minus the exercise price.

C exercise price minus the value of the underlying.

27 For a European call option with two months until expiration, if the spot price is below the exercise price, the call option will *most likely* have:

A zero time value.

B positive time value.

C positive exercise value.

28 When the price of the underlying is below the exercise price, a put option is:

A in-the-money.

B at-the-money.

C out-of-the-money.

29 If the risk-free rate increases, the value of an in-the-money European put option will *most likely*:

A decrease.

B remain the same.

C increase.

30 The value of a European call option is inversely related to the:

A exercise price.

B time to expiration.

C volatility of the underlying.

31 The table below shows three European call options on the same underlying:

	Time to Expiration	Exercise Price
Option 1	3 months	$100
Option 2	6 months	$100
Option 3	6 months	$105

The option with the highest value is *most likely*:

A Option 1.

B Option 2.

C Option 3.

32 The value of a European put option can be either directly or inversely related to the:

A exercise price.

B time to expiration.

C volatility of the underlying.

33 Prior to expiration, the lowest value of a European put option is the greater of zero or the:

A exercise price minus the value of the underlying.

B present value of the exercise price minus the value of the underlying.

C value of the underlying minus the present value of the exercise price.

34 A European put option on a dividend-paying stock is *most likely* to increase if there is an increase in:

A carrying costs.

B the risk-free rate.

C dividend payments.

35 Based on put-call parity, a trader who combines a long asset, a long put, and a short call will create a synthetic:

A long bond.

B fiduciary call.

C protective put.

36 Which of the following transactions is the equivalent of a synthetic long call position?

A Long asset, long put, short call

B Long asset, long put, short bond

C Short asset, long call, long bond

37 Which of the following is *least likely* to be required by the binomial option pricing model?

- **A** Spot price
- **B** Two possible prices one period later
- **C** Actual probabilities of the up and down moves

38 To determine the price of an option today, the binomial model requires:

- **A** selling one put and buying one offsetting call.
- **B** buying one unit of the underlying and selling one matching call.
- **C** using the risk-free rate to determine the required number of units of the underlying.

39 Assume a call option's strike price is initially equal to the price of its underlying asset. Based on the binomial model, if the volatility of the underlying decreases, the lower of the two potential payoff values of the hedge portfolio:

- **A** decreases.
- **B** remains the same.
- **C** increases.

40 Based on the binomial model, an increase in the actual probability of an upward move in the underlying will result in the option price:

- **A** decreasing.
- **B** remaining the same.
- **C** increasing.

41 If a call option is priced higher than the binomial model predicts, investors can earn a return in excess of the risk-free rate by:

- **A** investing at the risk-free rate, selling a call, and selling the underlying.
- **B** borrowing at the risk-free rate, buying a call, and buying the underlying.
- **C** borrowing at the risk-free rate, selling a call, and buying the underlying.

42 An at-the-money American call option on a stock that pays no dividends has three months remaining until expiration. The market value of the option will *most likely* be:

- **A** less than its exercise value.
- **B** equal to its exercise value.
- **C** greater than its exercise value.

43 At expiration, American call options are worth:

- **A** less than European call options.
- **B** the same as European call options.
- **C** more than European call options.

44 Which of the following circumstances will *most likely* affect the value of an American call option relative to a European call option?

- **A** Dividends are declared
- **B** Expiration date occurs
- **C** The risk-free rate changes

45 Combining a protective put with a forward contract generates equivalent outcomes at expiration to those of a:

- **A** fiduciary call.
- **B** long call combined with a short asset.
- **C** forward contract combined with a risk-free bond.

46 Holding an asset and buying a put on that asset is equivalent to:

A initiating a fiduciary call.

B buying a risk-free zero-coupon bond and selling a call option.

C selling a risk-free zero-coupon bond and buying a call option.

47 If an underlying asset's price is less than a related option's strike price at expiration, a protective put position on that asset versus a fiduciary call position has a value that is:

A lower.

B the same.

C higher.

48 Based on put–call parity, which of the following combinations results in a synthetic long asset position?

A A long call, a short put, and a long bond

B A short call, a long put, and a short bond

C A long call, a short asset, and a long bond

49 For a holder of a European option, put–call–forward parity is based on the assumption that:

A no arbitrage is possible within the spot, forward, and option markets.

B the value of a European put at expiration is the greater of zero or the underlying value minus the exercise price.

C the value of a European call at expiration is the greater of zero or the exercise price minus the value of the underlying.

50 Under put–call–forward parity, which of the following transactions is risk free?

A Short call, long put, long forward contract, long risk-free bond

B Long call, short put, long forward contract, short risk-free bond

C Long call, long put, short forward contract, short risk-free bond

SOLUTIONS

1 C is correct. An asset's current price, S_0, is determined by discounting the expected future price of the asset by r (the risk free rate) plus λ (the risk premium) over the period from 0 to T, as illustrated in the following equation:

$$S_0 = \frac{E(S_T)}{(1 + r + \lambda)^T}$$

Thus, an asset's current price inversely relates to its level of risk via the related risk premium, λ.

A is incorrect because an asset's current price in spot markets is calculated using the risk-free rate plus a risk premium.

B is incorrect because an asset's current price in spot markets is inversely related, not directly related, to its level of risk.

2 A is correct. An illiquid position is a limit to arbitrage because it may be difficult to realize gains of an illiquid offsetting position. A significant opportunity arises from a sufficiently large price differential or a small price differential that can be employed on a very large scale.

3 A is correct. Some arbitrage opportunities represent such small price discrepancies that they are only worth exploiting if the transaction costs are low. An arbitrage opportunity may require short-selling assets at costs that eliminate any profit potential. If the law of one price holds, there is no arbitrage opportunity.

4 C is correct. Arbitrage is a type of transaction undertaken when two assets or portfolios produce identical results but sell for different prices. A trader buys the asset or portfolio with the lower price and sells the asset or portfolio with the higher price, generating a net inflow of funds at the start of the holding period. Because the two assets or portfolios produce identical results, a long position in one and short position in the other means that at the end of the holding period, the payoffs offset. Therefore, there is no money gained or lost at the end of the holding period, so there is no risk.

5 B is correct. A long asset and a short risk-free asset (meaning to borrow at the risk-free rate) can be combined to produce a long derivative position.

A is incorrect because a short derivative and a long asset combine to produce a position equivalent to a long risk-free bond, not a long derivative.

C is incorrect because a short derivative and a short risk-free bond combine to produce a position equivalent to a short asset, not a long derivative.

6 B is correct. Virtually all derivative pricing models discount the expected payoff of the derivative at the risk-free rate.

A is incorrect because derivatives are priced by assuming that the market is free of arbitrage opportunities via the principle of no arbitrage, not by assuming that the market offers them.

C is incorrect because the application of a risk premium to the expected payoff of the derivative and its risk is not appropriate in the pricing of derivatives. An investor's risk premium is not relevant to pricing a derivative.

7 B is correct. The forward price is agreed upon at the start of the contract and is the fixed price at which the underlying will be purchased (or sold) at expiration. Payment is made at expiration. The value of the forward contract may change over time, but the forward price does not change.

8 C is correct. The price of a forward contract is a contractually fixed price, established at initiation, at which the underlying will be purchased (or sold) at expiration. The value of a forward contract at initiation is zero; therefore, the forward price is greater than the value of the forward contract at initiation.

9 B is correct. The value of the forward contract, unlike its price, will adjust as market conditions change. The forward price is fixed at initiation.

10 A is correct. When a forward contract expires, if the spot price is higher than the forward price, the long party profits from paying the lower forward price for the underlying. Therefore, the forward contract has a positive value to the long party and a negative value to the short party. However, if the forward price is higher than the spot price, the short party profits from receiving the higher forward price (the contract value is positive to the short party and negative to the long party).

11 B is correct. At initiation, the forward price is the future value of the spot price (spot price compounded at the risk-free rate over the life of the contract). If the forward price were set to the spot price or the present value of the spot price, it would be possible for one side to earn an arbitrage profit by selling the asset and investing the proceeds until contract expiration.

12 A is correct. The forward price of each stock is found by compounding the spot price by the risk-free rate for the period and then subtracting the future value of any benefits and adding the future value of any costs. In the absence of any benefits or costs, the one-year forward prices of BWQ and ZER should be equal. After subtracting the benefits related to BWQ, the one-year forward price of BWQ is lower than the one-year forward price of ZER.

13 A is correct. An asset's forward price is increased by the future value of any costs and decreased by the future value of any benefits: $F_0(T) = S_0(1 + r)^T - (\gamma - \theta)(1 + r)^T$. If the net cost of carry (benefits less costs) is positive, the forward price is lower than if the net cost of carry was zero.

14 C is correct. When a commodity's storage costs exceed its convenience yield benefits, the net cost of carry (benefits less costs) is negative. Subtracting this negative amount from the spot price compounded at the risk-free rate results in an addition to the compounded spot price. The result is a commodity forward price which is higher than the spot price compounded. The commodity's forward price is less than the spot price compounded when the convenience yield benefits exceed the storage costs and the commodity's forward price is the same as the spot price compounded when the costs equal the benefits.

15 C is correct. The convenience yield is a benefit of holding the asset and generally exists when a commodity is in short supply. The future value of the convenience yield is subtracted from the compounded spot price and reduces the commodity's forward price relative to it spot price. The opportunity cost is the risk-free rate. In the absence of carry costs, the forward price is the spot price compounded at the risk-free rate and will exceed the spot price. Dividends are benefits that reduce the forward price but the lack of dividends has no effect on the spot price relative to the forward price of a commodity in short supply.

16 B is correct. When interest rates are constant, forwards and futures will likely have the same prices. The price differential will vary with the volatility of interest rates. In addition, if futures prices and interest rates are uncorrelated, forward and futures prices will be the same. If futures prices are positively correlated with interest rates, futures contracts are more desirable to holders of long positions than are forwards. This is because rising prices lead to future profits that are reinvested in periods of rising interest rates, and falling prices lead to losses that occur in periods of falling interest rates. If futures prices are

negatively correlated with interest rates, futures contracts are less desirable to holders of long positions than are forwards. The more desirable contract will tend to have the higher price.

17 C is correct. Futures contracts are marked-to-market on a daily basis. The accumulated gains and losses from the previous day's trading session are deducted from the accounts of those holding losing positions and transferred to the accounts of those holding winning positions. Futures contracts trade on an exchange, forward contracts are over-the-counter transactions. Typically both forward and futures contracts are initiated at a zero value.

18 A is correct. If futures prices and interest rates are negatively correlated, forwards are more desirable to holders of long positions than are futures. This is because rising prices lead to futures profits that are reinvested in periods of falling interest rates. It is better to receive all of the cash at expiration under such conditions. If futures prices and interest rates are uncorrelated, forward and futures prices will be the same. If futures prices are positively correlated with interest rates, futures contracts are more desirable to holders of long positions than are forwards.

19 B is correct. Valuation of the swap during its life appeals to replication and the principle of arbitrage. Valuation consists of reproducing the remaining payments on the swap with other transactions. The value of that replication strategy is the value of the swap. The swap price is typically set such that the swap contract has a value of zero at initiation. The value of a swap contract will change during the life of the contract as the value of the underlying changes in value.

20 C is correct. Replication is the key to pricing a swap. The swap price is determined at initiation by replication. The value (not the price) of the swap is typically zero at initiation and the fixed swap price is typically determined such that the value of the swap will be zero at initiation.

21 B is correct. The principal of replication articulates that the valuation of a swap is the present value of all the net cash flow payments from the swap, not simply the present value of the fixed payments of the swap or the present value of the underlying at the end of the contract.

22 B is correct. When two parties engage in a series of forward contracts and initially agree on a price of $FS_0(T)$, some of the forward contracts have positive values and some have negative values, but their combined value equals zero.

A is incorrect because for a swap, all payments are fixed and equal, not variable.

C is incorrect because forward prices are determined by the spot price and the net cost of carry, meaning that forward contracts expiring at different times will have different prices, not the same price.

23 C is correct. On each payment date, the swap owner receives a payment based on the value of the underlying at the time of each respective payment.

A is incorrect because in a swap involving a series of fixed payments exchanged for a series of floating payments, each floating payment reflects the value of the underlying at the time of payment, not a designated value at contract initiation.

B is incorrect because in a swap involving a series of fixed payments exchanged for a series of floating payments, each floating payment is based on the value of the underlying at the time of each respective payment, not on the market value at the end of the swap.

24 B is correct. If the underlying has a value equal to the exercise price at expiration, both options will have zero value since they both have the same exercise price. For example, if the exercise price is \$25 and at expiration the underlying

price is \$25, both the call option and the put option will have a value of zero. The value of an option cannot fall below zero. The holder of an option is not obligated to exercise the option; therefore, the options each have a minimum value of zero. If the call has a positive value, the put, by definition, must have a zero value and vice versa. Both cannot have a positive value.

25 C is correct. A European put option will be valuable at expiration if the exercise price is greater than the underlying price. The holder can put (deliver) the underlying and receive the exercise price which is higher than the spot price. A European put option would be worthless if the exercise price was equal to or less than the underlying price.

26 B is correct. The value of a European call option at expiration is the greater of zero or the value of the underlying minus the exercise price.

27 B is correct. A European call option with two months until expiration will typically have positive time value, where time value reflects the value of the uncertainty that arises from the volatility in the underlying. The call option has a zero exercise value if the spot price is below the exercise price. The exercise value of a European call option is Max($0,S_t - X$), where S_t is the current spot price at time t and X is the exercise price.

28 A is correct. When the price of the underlying is below the exercise price for a put, the option is said to be in-the-money. If the price of the underlying is the same as the exercise price, the put is at-the-money and if it is above the exercise price, the put is out-of-the-money.

29 A is correct. An in-the-money European put option decreases in value with an increase in the risk-free rate. A higher risk-free rate reduces the present value of any proceeds received on exercise.

30 A is correct. The value of a European call option is inversely related to the exercise price. A lower exercise price means there are more potential outcomes at which the call expires in-the-money. The option value will be greater the lower the exercise price. For a higher exercise price, the opposite is true. Both the time to expiration and the volatility of the underlying are directly (positively) related to the value of a European call option.

31 B is correct. The value of a European call option is inversely related to the exercise price and directly related to the time to expiration. Option 1 and Option 2 have the same exercise price; however, Option 2 has a longer time to expiration. Consequently, Option 2 would likely have a higher value than Option 1. Option 2 and Option 3 have the same time to expiration; however, Option 2 has a lower exercise price. Thus, Option 2 would likely have a higher value than Option 3.

32 B is correct. The value of a European put option can be either directly or indirectly related to time to expiration. The direct effect is more common, but the inverse effect can prevail the longer the time to expiration, the higher the risk-free rate, and the deeper in-the-money is the put. The value of a European put option is directly related to the exercise price and the volatility of the underlying.

33 B is correct. Prior to expiration, the lowest value of a European put is the greater of zero or the present value of the exercise price minus the value of the underlying.

34 C is correct. Payments, such as dividends, reduce the value of the underlying which increases the value of a European put option. Carrying costs reduce the value of a European put option. An increase in the risk-free interest rate may decrease the value of a European put option.

35 A is correct. A long bond can be synthetically created by combining a long asset, a long put, and a short call. A fiduciary call is created by combining a long call with a risk free bond. A protective put is created by combining a long asset with a long put.

36 B is correct. According to put–call parity, a synthetic call can be constructed by combining a long asset, long put, and short bond positions.

37 C is correct. The actual probabilities of the up and down moves in the underlying do not appear in the binomial option pricing model, only the pseudo or "risk-neutral" probabilities. Both the spot price of the underlying and two possible prices one period later are required by the binomial option pricing model.

38 C is correct. Pricing an option relies on the facts that a perfectly hedged investment earns the risk-free rate and that, based on the binomial option pricing model, the size of the two possible changes in the option price (meaning the potential step up or step down in the option value) after one period are equivalent.

39 B is correct. When the volatility of the underlying decreases, the value of the option also decreases, meaning that the upper payoff value of the hedge portfolio combining them declines. However, the lower payoff value remains at zero.

40 B is correct. The binomial model does not consider the actual probabilities of upward and downward movements in determining the option value. Thus, a change in this probability has no effect on the calculated option price.

41 C is correct. If an option is trading above the value predicted by the binomial model, investors can engage in arbitrage by selling a call, buying shares of the underlying, and funding the transaction by borrowing at the risk-free rate. This will earn a return in excess of the risk-free rate.

42 C is correct. Prior to expiration, an American call option will typically have a value in the market that is greater than its exercise value. Although the American option is at-the-money and therefore has an exercise value of zero, the time value of the call option would likely lead to the option having a positive market value.

43 B is correct. At expiration, the values of American and European call options are effectively the same; both are worth the greater of zero and the exercise value.

44 A is correct. When a dividend is declared, an American call option will have a higher value than a European call option because an American call option holder can exercise early to capture the value of the dividend. At expiration, both types of call options are worth the greater of zero and the exercise value. A change in the risk-free rate does not affect the relative values of American and European call options.

45 A is correct. Put–call forward parity demonstrates that the outcome of a protective put with a forward contract (long put, long risk-free bond, long forward contract) equals the outcome of a fiduciary call (long call, long risk-free bond). The outcome of a protective put with a forward contract is also equal to the outcome of a protective put with asset (long put, long asset).

46 A is correct. Under put–call parity, initiating a fiduciary call (buying a call option on an asset that expires at time T together with a risk-free zero-coupon bond that also expires at time T) is equivalent to holding the same asset and initiating a protective put on it (buying a put option with an exercise price of X that can be used to sell the asset for X at time T).

47 B is correct. On the one hand, buying a call option on an asset and a risk-free bond with the same maturity is known as a fiduciary call. If the fiduciary call expires in the money (meaning that the value of the call, $S_T - X$, is greater than the risk-free bond's price at expiration, X), then the total value of the fiduciary call is $(S_T - X) + X$, or S_T. On the other hand, holding an underlying asset, S_T, and buying a put on that asset is known as a protective put. If the put expires out of the money, meaning that the value of the asset, S_T, is greater than the put's value at expiration, 0, then the total value of the protective put is $S_T - 0$, or S_T. A protective put and a fiduciary call produce the same result.

48 A is correct. One can synthetically create a long asset position by buying a call, shorting a put, and buying a bond.

B is incorrect because combining a short call and a short bond with the right to sell (not buy) another asset via a long put could not result in a new synthetic long asset position.

C is incorrect because combining a long call, a short asset, and a long bond creates a long put, not a synthetic long asset.

49 A is correct. Put–call–forward parity is based on the assumption that no arbitrage is possible within the spot, forward, and option markets.

B is incorrect because the value of a European put at expiration is the greater of either zero or the exercise price minus the value of the underlying, not the greater of zero or the underlying value minus the exercise price. In addition, put–call–forward parity is related to the equality of a fiduciary call and a synthetic protective put or to a protective put and a synthetic fiduciary call, not specifically to the value of a put at expiration.

C is incorrect because the value of a European call at expiration is the greater of either zero or the underlying value minus the exercise price, not the greater of zero or the exercise price minus the value of the underlying. In addition, put–call–forward parity is related to the equality of a fiduciary call and a synthetic protective put or to a protective put and a synthetic fiduciary call, not specifically to the value of a call at expiration.

50 A is correct. Purchasing a long forward contract and a risk-free bond creates a synthetic asset. Combining a long synthetic asset, a long put, and a short call is risk free because its payoffs produce a known cash flow of the value of the exercise price.

Alternative Investments

STUDY SESSION

Study Session 16	Alternative Investments

TOPIC LEVEL LEARNING OUTCOME

The candidate should be able to demonstrate a working knowledge of alternative investments, including hedge funds, private capital, real estate, natural resources, and infrastructure. The candidate should be able to describe key attributes and considerations in adding these investments to a portfolio.

Investors often turn to alternative investments for potential diversification benefits and higher returns. As a result, alternative investments now represent meaningful allocations in many institutional and private wealth portfolios. Although the category of "alternative investments" is not always clearly or precisely defined, alternative investments often have a number of characteristics in common. These include lower levels of liquidity, transparency, and disclosure vs. traditional asset classes (equity, fixed income), more complex legal structures, and performance-based compensation arrangements.

ALTERNATIVE INVESTMENTS

STUDY SESSION

16

Alternative Investments

This study session provides an overview of alternative investments, including hedge funds, private capital, real estate, natural resources, and infrastructure. Each is examined with emphasis on their distinguishing characteristics, potential benefits, and risks. Similarities and differences with traditional investments (stocks, bonds) are also considered.

READING ASSIGNMENTS

Reading 47	Introduction to Alternative Investments by Steve Balaban, CFA, Steven G. Bloom, CFA, David Burkart, CFA, Nasir Hasan, and Barclay T. Leib, CFE, CAIA

READING

47

Introduction to Alternative Investments

by Steve Balaban, CFA, Steven G. Bloom, CFA, David Burkart, CFA, Nasir Hasan, and Barclay T. Leib, CFE, CAIA

Steve Balaban, CFA, is at Mink Capital Inc. (Canada). Steven G. Bloom, CFA, is at ARC Fiduciary (USA). David Burkart, CFA, is at Coloma Capital Futures, LLC (USA). Nasir Hasan is at Ernst & Young (UAE). Barclay T. Leib, CFE, CAIA, is at and Spring Advisors LLC (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** describe types and categories of alternative investments;
☐	**b.** describe characteristics of direct investment, co-investment, and fund investment methods for alternative investments;
☐	**c.** describe investment and compensation structures commonly used in alternative investments;
☐	**d.** explain investment characteristics of hedge funds;
☐	**e.** explain investment characteristics of private capital;
☐	**f.** explain investment characteristics of natural resources;
☐	**g.** explain investment characteristics of real estate;
☐	**h.** explain investment characteristics of infrastructure;
☐	**i.** describe issues in performance appraisal of alternative investments;
☐	**j.** calculate and interpret returns of alternative investments on both before-fee and after-fee bases.

INTRODUCTION

1

a describe types and categories of alternative investments

In this section, we explain what alternative investments are and why assets under management in alternative investments have grown in recent decades. We also explain how alternative investments differ from traditional investments, and we examine their perceived investment merit. We conclude this section with a brief overview of the various categories of alternative investments; these categories will be explored further in later sections.

"Alternative investments" is a label for a disparate group of investments that are distinguished from long-only, publicly traded investments in stocks, bonds, and cash (often referred to as traditional investments). The terms "traditional" and "alternative" should not imply that alternatives are necessarily uncommon or that they are relatively recent additions to the investment universe. Alternative investments include such assets as real estate and commodities, which are arguably two of the oldest types of investments.

Alternative investments also include non-traditional approaches to investing within special vehicles, such as private equity funds and hedge funds. These funds may give the manager flexibility to use derivatives and leverage, to make investments in illiquid assets, and to take short positions. The assets in which these vehicles invest can include traditional assets (stocks, bonds, and cash) as well as less traditional assets. Management of alternative investments is typically active. Alternative investments often have many of the following characteristics:

- Narrow specialization of the investment managers
- Relatively low correlation of returns with those of traditional investments
- Less regulation and less transparency than traditional investments
- Limited historical risk and return data
- Unique legal and tax considerations
- Higher fees, often including performance or incentive fees
- Concentrated portfolios
- Restrictions on redemptions (i.e., "lockups" and "gates")

1.1 Why Investors Consider Alternative Investments

Assets under management in alternative investments have grown rapidly since the mid-1990s. This growth has largely occurred because of interest in these investments from institutions, such as endowments and pension funds, and from family offices seeking diversification and return opportunities. Alternative investments offer broader diversification (because of their lower correlation with traditional asset classes), opportunities for enhanced returns (by increasing the portfolio's risk–return profile), and potentially increased income through higher yields (particularly compared with traditional investments in low–interest rate periods).

The 2019 annual report for the Yale University endowment provides one institutional investor's reasoning behind the attractiveness of investing in alternatives:

> The heavy [75.2%] allocation to nontraditional asset classes stems from the diversifying power they provide to the portfolio as a whole. Alternative assets, by their very nature, tend to be less efficiently priced than traditional marketable securities, providing an opportunity to exploit market inefficiencies through active management. Today's portfolio has significantly higher expected returns and lower volatility than the 1989 portfolio.[1]

This quote neatly illustrates the expected characteristics of alternative investments: diversifying power (low correlations between returns), higher expected returns (positive absolute return), and illiquid and potentially less efficient markets. These links

---

1 Yale University, "Financial Report 2018–2019 Yale University" (2019, p. 18). https://your.yale.edu/sites/default/files/annual-report-2018-2019.pdf (accessed 27 April 2020).

also highlight the importance of having the ability and willingness to take a long-term perspective. Allocating a portion of an endowment portfolio to alternative investments is not unique to Yale. INSEAD, as of April 2020, had allocated 33% of its endowment to private market strategies, including private equity, private debt, and real estate.[2] These examples are not meant to imply that every university endowment fund invests in alternative investments, but many do.

Family offices have also embraced alternative investments. According to the Global Family Office Report 2019 from UBS, over 40% of the average family office portfolio was invested in alternative assets.[3] According to Willis Towers Watson's "Global Pension Assets Study 2020," the typical pension plan had a 23% allocation to alternative investments at the end of 2019.[4]

Alternative investments are not free of risk, of course, and their returns may be correlated with those of other investments, especially in periods of financial crisis. Over a long historical period, the average correlation of returns from alternative investments with those of traditional investments may be low, but in any particular period, the correlation can differ from the average. During periods of economic crisis, correlations among many assets (both alternative and traditional) can increase dramatically.

KNOWLEDGE CHECK

True or false: Alternative investments focus solely on the private markets.

A True

B False

Solution:

B is correct. Although many alternative investments are focused on private markets, there are alternative investments, such as hedge funds, that focus on the public markets.

1.2 Categories of Alternative Investments

Considering the variety of alternative investments, it is not surprising that no consensus exists on a definitive list. There is even considerable debate around categories versus sub-categories. The following list offers one approach to broad category definitions, and each category is described in detail later in this reading.

- **Hedge funds.** Hedge funds are private investment vehicles that manage portfolios of securities and/or derivative positions using a variety of strategies. Although hedge funds may be invested entirely in traditional assets, these vehicles are considered alternative because of their private nature. Hedge funds typically have more leeway to pursue investments and strategies offering the potential for higher returns, whether absolute or compared with a specific market benchmark, but these strategies may increase the risk of investment loss. They may involve long and short positions and may be highly leveraged. Some aim to deliver investment performance that is independent of broader market performance.

2 INSEAD, "Finances and Endowment" (sec. 3), in "INSEAD Annual Report 2018/2019" (2019). https://annual-report.insead.edu/finances-and-endowment (accessed 27 April 2020).

3 Campden Wealth Limited and UBS, "The Global Family Office Report 2019" (2019, p. 17). www.ubs.com/global/en/wealth-management/uhnw/global-family-office-report/global-family-office-report-2019.html.

4 Thinking Ahead Institute, "Global Pension Assets Study 2020," Willis Towers Watson (2020).

- **Private Capital**
  - **Private equity.** Investors participate in private equity through direct investments or indirectly through **private equity funds**. Private equity funds generally invest in companies, whether startups or established firms, that are not listed on a public exchange, or they invest in public companies with the intent to take them private. The majority of private equity activity involves leveraged buyouts of established profitable and cash-generating companies with solid customer bases, proven products, and high-quality management. **Venture capital funds**, a specialized form of private equity that typically involves investing in or providing financing to startup or early-stage companies with high growth potential, represent a small portion of the private equity market.
  - **Private debt.** Private debt largely encompasses debt provided to private entities. Forms of private debt include direct lending (private loans with no intermediary), mezzanine loans (private subordinated debt), venture debt (private loans to startup or early-stage companies that may have little or negative cash flow), and distressed debt (debt extended to companies that are "distressed" because of such issues as bankruptcy or other complications with meeting debt obligations).

KNOWLEDGE CHECK

Describe what the majority of private equity activity includes.

Solution:

The majority of private equity activity involves leveraged buyouts of established profitable and cash-generating companies with solid customer bases, proven products, and high-quality management.

- **Real estate.** Real estate investments are made in buildings or land, either directly or indirectly. The growing popularity of securitizations broadened the definition of real estate investing. It now includes private commercial real estate equity (e.g., ownership of an office building), private commercial real estate debt (e.g., directly issued loans or mortgages on commercial property), public real estate equity (e.g., real estate investment trusts, or REITs), and public real estate debt (e.g., mortgage-backed securities).
- **Natural Resources**
  - **Commodities.** Commodity investments may take place in physical commodity products, such as grains, metals, and crude oil, either through owning cash instruments, using derivative products, or investing in businesses engaged in the production of physical commodities. The main vehicles investors use to gain exposure to commodities are commodity futures contracts and funds benchmarked to commodity indexes. Commodity indexes are typically based on various underlying commodity futures.
  - **Agricultural land (or farmland).** Agricultural land is for the cultivation of livestock or plants, and agricultural land investing covers various strategies, including the purchase of farmland in order to lease it back to farmers or receive a stream of income from the growth, harvest, and sale of crops (e.g., corn, cotton, wheat) or livestock (e.g., cattle).

- **Timberland.** Investing in timberland generally involves investing capital in natural forests or managed tree plantations in order to earn a return when the trees are harvested. Timberland investors often rely on various drivers, such as biological growth, to increase the value of the trees so the wood can be sold at favorable prices in the future.

- **Infrastructure.** Infrastructure assets are capital-intensive, long-lived real assets, such as roads, dams, and schools, that are intended for public use and provide essential services. Infrastructure assets may be financed, owned, and operated by governments, but private sector investment is on the rise. An increasingly common approach to infrastructure investing is a **public–private partnership (PPP)** approach, in which governments and investors each have a stake. Infrastructure investments provide exposure to asset cash flows, but the asset itself is generally part of a long-term concession agreement, ultimately going back to the public authority. Investors may gain exposure to these assets directly or indirectly. Indirect investment vehicles include shares of companies, exchange-traded funds (ETFs), private equity funds, listed funds, and unlisted funds that invest in infrastructure.
- **Other.** Other alternative investments may include tangible assets, such as fine wine, art, antique furniture and automobiles, stamps, coins, and other collectibles, and intangible assets, such as patents and litigation actions.

PRACTICE QUESTIONS FOR LOS A

Question 1

Which of the following is *least likely* to be considered an alternative investment?

A Real estate

B Commodities

C Long-only equity funds

Solution:

C is correct. Long-only equity funds are typically considered traditional investments, and real estate and commodities are typically classified as alternative investments.

Question 2

An investor is seeking an investment that can take long and short positions, may use multi-strategies, and historically exhibits low correlation with a traditional investment portfolio. The investor's goals will be *best* satisfied with an investment in:

A real estate.

B a hedge fund.

C a private equity fund.

Solution:

B is correct. Hedge funds may use a variety of strategies, generally have a low correlation with traditional investments, and may take long and short positions.

Question 3

Relative to traditional investments, alternative investments are *least likely* to be characterized by:

A high levels of transparency.

B limited historical return data.

C significant restrictions on redemptions.

Solution:

A is correct. Alternative investments are characterized as typically having low levels of transparency.

Question 4

Alternative investment funds are typically managed:

A actively.

B to generate positive beta return.

C assuming that markets are efficient.

Solution:

A is correct. There are many approaches to managing alternative investment funds, but typically these funds are actively managed.

Question 5

Compared with traditional investments, alternative investments are *more likely* to have:

A greater use of leverage.

B long-only positions in liquid assets.

C more transparent and reliable risk and return data.

Solution:

A is correct. Investing in alternative investments is often pursued through such special vehicles as hedge funds and private equity funds, which have flexibility to use leverage. Alternative investments include investments in such assets as real estate, which is an illiquid asset, and investments in such special vehicles as private equity and hedge funds, which may make investments in illiquid assets and take short positions. Obtaining information on strategies used and identifying reliable measures of risk and return are challenges of investing in alternatives.

Question 6

The potential benefits of allocating a portion of a portfolio to alternative investments include:

A ease of manager selection.

B improvement in the portfolio's risk–return relationship.

C accessible and reliable measures of risk and return.

Solution:

B is correct. Adding alternative investments to a portfolio may provide diversification benefits because of these investments' less-than-perfect correlation with other assets in the portfolio. As a result, allocating a portion of one's funds to alternatives could potentially result in an improved risk–return relationship.

Challenges to allocating a portion of a portfolio to alternative investments include obtaining reliable measures of risk and return and selecting portfolio managers for the alternative investments.

INVESTMENT METHODS

2

b describe characteristics of direct investment, co-investment, and fund investment methods for alternative investments

This section introduces three methods of investing in alternative investments: direct investing, co-investing, and fund investing. Their advantages and disadvantages are examined, and a brief discussion of due diligence concludes the section.

2.1 Methods of Investing in Alternative Investments

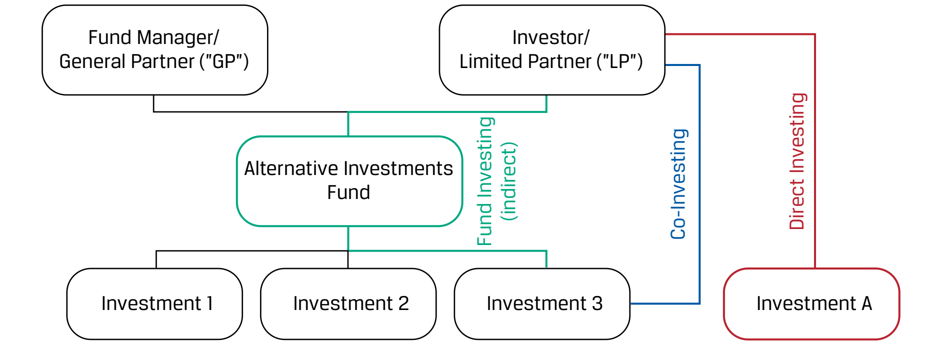
In **fund investing**, the investor contributes capital to a fund, and the fund identifies, selects, and makes investments on the investor's behalf. For the fund's services, the investor is charged a fee based on the amount of the assets being managed, and a performance fee is applied if the fund manager delivers superior results. Fund investing can be viewed as an indirect method of investing in alternative assets. Fund investors have little or no leeway in the sense that their investment decisions are limited to either investing in the fund or not. Furthermore, fund investors are typically unable to affect the fund's underlying investments. Note that fund investing is available for all major alternative investment types, including hedge funds, private capital, real estate, infrastructure, and natural resources.

In **co-investing**, the investor invests in assets *indirectly* through the fund but also possesses rights (known as co-investment rights) to invest *directly* in the same assets. Through co-investing, an investor is able to make an investment *alongside* a fund when the fund identifies deals; the investor is not limited to participating in the deal solely by investing in the fund. Exhibit 1 illustrates the co-investing method: The investor invests in one deal (labeled "Investment #3") indirectly via fund investing while investing an additional amount directly via a co-investment.

Direct investing occurs when an investor makes a direct investment in an asset (labeled "Investment A") without the use of an intermediary. In private equity, this may mean the investor purchases a direct stake in a private company without the use of a special vehicle, such as a fund. Direct investors have great flexibility and control when it comes to choosing their investments, selecting their preferred methods of financing, and planning their approach. The direct method of investing in alternative assets is typically reserved for larger and more sophisticated investors and usually applies to private equity and real estate. Sizable investors, such as major pensions and sovereign wealth funds, however, may also invest directly in infrastructure and natural resources.

Institutional investors typically begin investing in alternative investments via funds. Then, as they gain experience, they can begin to invest via co-investing and direct investing. The largest and most sophisticated direct investors (such as some sovereign wealth funds) compete with fund managers for access to the best investment opportunities.

Exhibit 1 Diagram of Co-Investing in Alternative Assets



KNOWLEDGE CHECK

True or false: In co-investing, the investor is able to invest both directly and indirectly in the same assets.

A True

B False

Solution:

A is correct. In co-investing, the investor is able to invest both directly and indirectly in the same assets.

2.2 Advantages and Disadvantages of Direct Investing, Co-Investing, and Fund Investing

The direct investing, co-investing, and fund investing approaches naturally have distinct advantages and disadvantages.

2.2.1 *Advantages of Fund Investing*

The primary advantages of fund investing include the professional services offered by fund managers, a lower level of investor involvement (compared with the direct and co-investing methods), and access to alternative investments without the prerequisite of advanced expertise. Additionally, diversification benefits stem from the multiple investments found in a single investment vehicle overseen by a fund manager. (Consider instead the time and attention that would be required of a single investor to achieve similar diversification levels by directly investing in and managing multiple alternative assets.) Fund investing in alternative assets demands less participation from the investor than the direct and co-investing approaches, because it is up to the fund manager to identify, select, and manage the fund's investments. The investor may also heavily rely on the fund manager to carry out the due diligence that accompanies the investment process. Finally, specialist fund managers may expand the investment universe for an investor who lacks expertise for the sector in question.

2.2.2 *Disadvantages of Fund Investing*

Fund investing in alternative investments is costly because the investor is required to pay management fees and performance fees that are typically higher than fees for traditional asset classes. The higher fees can be attributed to the increased costs of

(1) running an alternative fund that may have a more complex strategy or require more skills and resources to source, (2) conducting due diligence, and (3) managing opportunities that have limited publicly available information. Alternative managers may also command higher fees if they face little competition. Despite the fact that a fund may provide due diligence expertise when it comes to choosing investments, an investor is still required to conduct thorough due diligence when selecting the right fund in the first place. Selecting the right fund itself is not an easy task because of asymmetry of information.

2.2.3 *Advantages of Co-Investing*

Co-investors, who are essentially engaging in a hybrid of direct investing and fund investing, can learn from the fund's process and leverage the experience they gain to become better at direct investing. Many institutional investors begin with fund investing, because it provides instant diversification while requiring less costly minimum participation levels (lower minimums) and less due diligence. As investors' investment experience and investable assets grow, they may be granted "co-investment rights" with an investment in a fund, giving them a taste of direct investing. It works as follows: Alongside the fund's direct investment, investors co-invest an additional amount into that same investment often without paying management fees on the capital they used for the direct investment (a co-investment, in this case). Compared with fund investing, co-investing allows investors to be more actively involved with the management of their portfolio. Later, as their experience and capital increase, investors may prefer direct investing, bringing the expertise in-house and avoiding fund management fees.

2.2.4 *Disadvantages of Co-Investing*

Co-investors have reduced control over the investment selection process (compared with direct investing) and may be subject to adverse selection bias, where the fund manager makes less attractive investment opportunities available to the co-investor while allocating its own capital to more appealing deals. Co-investing, furthermore, requires more active involvement from the investor, who must evaluate both investment opportunities and the fund manager, which demands more resources, concentration, and expertise. Moreover, co-investors usually have a limited amount of time to make the decision to invest or not. The co-investing approach can prove to be challenging for smaller firms with limited resources and due diligence experience.

2.2.5 *Advantages of Direct Investing*

When an investor chooses direct investing, she avoids paying ongoing management fees to an external manager because the investor conducts the investment process on her own, bypassing the use of a special vehicle, such as the fund. Direct investing allows the investor to build a portfolio of investments to her exact requirements. Direct investing provides the greatest amount of flexibility for the investor and grants the highest level of control over how the asset is managed. For example, an investor who directly purchases an ownership stake in a business typically has the ability to influence important matters, such as selecting the management team and controlling the strategic direction and investment decisions of the company.

2.2.6 *Disadvantages of Direct Investing*

Compared with fund investing and co-investing, direct investing demands more investment expertise and a higher level of financial sophistication. Such experience would otherwise be provided by a fund manager or investment firm during the complex investment and management process. Furthermore, concentration increases risk: The direct investor won't enjoy the ready diversification benefits of fund investing; it would require time and resources for an investor to mirror this advantage through direct investing. Plus, fund managers may enjoy reputational benefits that see them secure participation in attractive investments unavailable to certain direct investors operating on their own behalf.

Exhibit 2 provides a summary of the advantages and disadvantages of the three methods of investing in alternative investments.

Exhibit 2 Summary of Advantages and Disadvantages of Fund Investing, Co-Investing, and Direct Investing

	Advantages	Disadvantages
Fund investing	■ Fund managers offer investment services and expertise ■ Lower level of investor involvement compared with the direct and co-investing methods ■ Access to alternative investments without possessing a high degree of investment expertise ■ Potentially valuable diversification benefits ■ Lower minimum capital requirements	■ Costly management and performance fees ■ Investor must conduct thorough due diligence when selecting the right fund because of the wide dispersion of fund manager returns
Co-investing	■ Investors can learn from the fund's process to become better at direct investing ■ Reduced management fees ■ Allows more active management of the portfolio compared with fund investing and allows for a deeper relationship with the manager	■ Reduced control over the investment selection process compared with direct investing ■ May be subject to adverse selection bias ■ Requires more active involvement compared with fund investing, which can be challenging if resources and due diligence experience are limited
Direct investing	■ Avoids paying ongoing management fees to an external manager ■ Greatest amount of flexibility for the investor ■ Highest level of control over how the asset is managed	■ Requires more investment expertise and a higher level of financial sophistication compared with fund investing and co-investing, resulting in higher internal investment costs ■ Less access to a fund's ready diversification benefits or the fund manager's sourcing network ■ Requires greater levels of due diligence because of the absence of a fund manager ■ Higher minimum capital requirements

KNOWLEDGE CHECK

Identify the advantages of direct investing for an investor.

Solution:

An advantage of direct investing is that the investor can avoid paying ongoing management fees. In addition, a direct investor has great flexibility and control when it comes to choosing his investments, selecting his preferred methods of financing, and planning his approach.

KNOWLEDGE CHECK

In direct investing, an investor puts capital in an asset or business:

A using a special purpose vehicle, such as a fund.

B using a separate business entity, such as a joint venture.

C without the use of an intermediary.

Solution:

C is correct. Direct investing occurs when an investor makes a direct investment in an asset without the use of an intermediary.

2.3 Due Diligence for Fund Investing, Direct Investing, and Co-Investing

One basic question an investor must consider when investing in alternative investments is whether to rely on the expertise of the fund manager or to undertake the investment selection process themselves via direct investing.

For direct investing, the investor has control over the choice of which company to invest in; however, the due diligence process of making an investment in a private company requires considerable expertise. In contrast, funds offer diversified portfolios, due diligence expertise, and the negotiation of favorable redemption terms, but these benefits come with additional fees. And although funds offer due diligence on the underlying investments, responsibility still falls to the investor for conducting due diligence when choosing among funds to invest in.

With co-investing, the investor is currently investing in a fund but is given an opportunity by the fund general partner (co-investment rights) to make an additional investment in a portfolio company (as shown in Exhibit 1). In this case, the investor will conduct direct due diligence on the portfolio company with the support of the general partner. A detailed discussion on due diligence is beyond the scope of this reading (especially for direct investing). However, this section provides a brief overview of the due diligence process.

2.3.1 *Due Diligence for Fund Investing*

Manager selection is a critical factor in portfolio performance. A manager should have a verifiable track record and display a high level of expertise and experience with the asset type. The asset management team should be assigned an appropriate workload and provided sufficient resources. Moreover, it should be rewarded with an effective compensation package to ensure alignment of interest, continuity, motivation, and thoughtful oversight of assets.

Fraud, although infrequent, is always a possibility. The investor should be skeptical of unusually good and overly consistent reported performance. Third-party custody of assets and independent verification of results can help reduce the chance of an investor being defrauded. Diversification across managers is also wise. Finally, separate accounts make theft more difficult because the investor retains custody of the assets and sometimes can select the prime broker or other service providers, binding them to the client's interest.

For an investor considering a new investment, a proper due diligence process should be carried out to ensure that the targeted investment is in compliance with its prospectus and that it will meet her investment strategy, risk and return objectives, and restrictions. Existing investors should monitor results and fund holdings to determine whether a fund has performed in line with expectations and continues to comply with its prospectus.

Exhibit 3 lists key items that should be considered in a typical fund due diligence process.

Exhibit 3 A Typical Due Diligence Process

Organization	■ Experience and quality of management team, compensation, and staffing ■ Analysis of prior and current funds ■ Track record/alignment of interests ■ Reputation and quality of third-party service providers (e.g., lawyers, auditors, prime brokers)
Portfolio management	■ Investment process ■ Target markets/asset types/strategies ■ Sourcing of investments ■ Role of operating partners ■ Underwriting ■ Environmental and engineering review process ■ Integration of asset management/acquisitions/dispositions ■ Disposition process, including its initiation and execution
Operations and controls	■ Reporting and accounting methodology ■ Audited financial statements and other internal controls ■ Valuations—frequency and approach(es) ■ Insurance and contingency plans
Risk management	■ Fund policies and limits ■ Risk management policy ■ Portfolio risk and key risk factors ■ Leverage and currency—risks/constraints/hedging
Legal review	■ Fund structure ■ Registrations ■ Existing/prior litigation
Fund terms	■ Fees (management and performance) and expenses ■ Contractual terms ■ Investment period and fund term and extensions ■ Carried interest ■ Distributions ■ Conflicts ■ Limited partners' rights ■ "Key person" and/or other termination procedures

2.3.2 *Due Diligence for Direct Investing*

Due diligence for direct investing requires the investor to conduct a thorough investigation into the target asset or business, including but not limited to the quality of its management team, the quality of its customers, the competitive landscape, revenue generation, risks, and so on. When considering direct investments in private equity, the investor conducts reference checks and interviews to evaluate the quality of interactions between the business and each of its stakeholders (e.g., customers and suppliers). Due diligence for private debt investing could entail credit analysis of borrowers to assess their ability to service the regular interest and principal payments of debt, and background checks on the owners and management team would round out the effort. In direct real estate, the building's occupancy rate and the quality of its structure and tenants should be analyzed prior to investing. In infrastructure, an investor would perform an assessment of the quality of the assets held and operated (e.g., an airport) and their ability to generate future cash flows. In conducting due diligence, direct investors often supplement their due diligence with analysis prepared by external consultants.

2.3.3 *Due Diligence for Co-Investing*

Given that direct investing is an element of co-investing, aspects of the due diligence process apply to both. In co-investing, investors often rely heavily on the due diligence conducted by the fund manager. Direct investing opportunities, however, are often sourced differently from co-investing opportunities, and so the level of independence sitting behind the due diligence can differ. Consider that direct investing opportunities are usually sourced by the direct investment team at a large pension or sovereign wealth fund, whereas co-investing opportunities are usually provided by the private equity, real estate, or infrastructure fund manager for the investor's consideration. Direct investing due diligence may be more independent than co-investing due diligence because the direct investing team is typically introduced to opportunities by third parties and they have more control over the due diligence process.

KNOWLEDGE CHECK

An investor with limited investment and due diligence experience will *most likely* invest in alternative assets using which method?

A Fund investing

B Co-investing

C Direct investing

Solution:

A is correct. An investor with limited investment and due diligence experience will likely choose fund investing to benefit from the expertise that a fund manager would provide.

PRACTICE QUESTIONS FOR LOS B

Question 1

From the perspective of the investor, the *most* active approach to investing in alternative investments is:

A co-investing.

B fund investing.

C direct investing.

Solution:

C is correct. From the perspective of the investor, direct investing is the most active approach to investing because of the absence of fund managers and the services and expertise they generally provide.

A is incorrect because co-investing includes fund investing, which requires less due diligence compared with direct investing.

B is incorrect because fund investing in alternative assets demands less participation from the investor compared with the direct and co-investing approaches because an investor depends on the fund manager to identify, select, and manage the fund's investments.

Question 2

In comparison to other alternative investment approaches, co-investing is *most likely*:

A more expensive.

B subject to adverse selection bias.

C the most flexible approach for the investor.

Solution:

B is correct. Co-investing may be subject to adverse selection bias. For example, the fund manager may make less attractive investment opportunities available to the co-investor while allocating its own capital to more appealing deals.

A is incorrect because co-investing is likely not more expensive than fund investing since co-investors can co-invest an additional amount alongside the fund directly in a fund investment without paying management fees on the capital that has been directly invested.

C is incorrect because direct investing, not co-investing, provides the greatest amount of flexibility for the investor.

Question 3

Relative to co-investing, direct investing due diligence is *most likely*:

A harder to control.

B more independent.

C equally thorough.

Solution:

B is correct. Direct investing due diligence may be more independent than that of co-investing because the direct investing team is typically introduced to opportunities by third parties rather than fund managers, as is customary in co-investing.

A is incorrect because the direct investing team has more control over the due diligence process compared with co-investing.

C is incorrect because due diligence for direct investing requires the investor to conduct a thorough investigation into the important aspects of a target asset or business, whereas in co-investing, fund managers typically provide investors with access to a data room so they can view the due diligence completed by the fund managers.

Question 4

The investment method that typically requires the greatest amount of or most thorough due diligence from an investor is:

A fund investing.

B co-investing.

C direct investing.

Solution:

C is correct. Due diligence in direct investing will usually be more thorough and more rigid from an investor's perspective because of the absence of a fund manager that would otherwise conduct a large portion of the necessary due diligence.

INVESTMENT AND COMPENSATION STRUCTURES

3

c describe investment and compensation structures commonly used in alternative investments

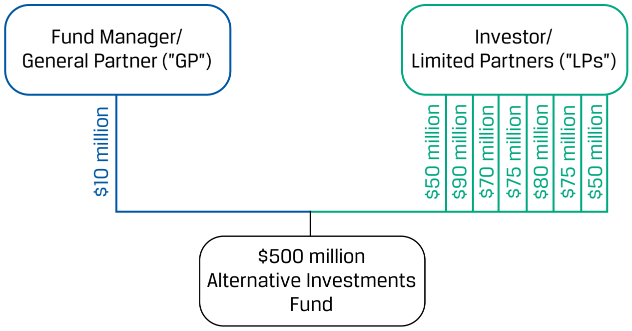
This section explores the investment partnership and compensation structures used in alternative investments. We examine the contractual provisions typically included in partnership agreements and why they exist to protect investors. An overview of the most common investment clauses is provided.

3.1 Partnership Structures

In the world of alternative investments, partnership structures are common. In limited partnerships, the fund manager is the **general partner (GP)** and investors are the **limited partners** (LPs). LPs, who are generally accredited investors (owing to legal restrictions on the fund), are expected to understand and be able to assume the risks associated with the investments, which are less regulated than offerings to the general public. The GP runs the business and theoretically bears unlimited liability for anything that goes wrong. The GP may also run multiple funds at a time.

Limited partners are outside investors who own a fractional interest in the partnership based on the amount of their initial investment and the terms set out in the partnership documentation. LPs commit to future investments, and the upfront cash outflow can be a small portion of their total commitment to the fund. Funds set up as limited partnerships typically have a limit on the number of LPs allowed to invest in the fund. LPs play passive roles and are not involved with the management of the fund (although co-investment rights allow for the LPs to make additional direct investments in the portfolio companies); the operations and decisions of the fund are controlled solely by the GP. Exhibit 4 illustrates the limited partnership structure for a hypothetical $500 million investment fund.

Exhibit 4 Example of a Limited Partnership Structure



The partnership between the GP and LPs is governed by a **limited partnership agreement (LPA)**, a legal document that outlines the rules of the partnership and establishes the framework that ultimately guides the fund's operations throughout its life. LPAs vary in length and complexity and may be dense with provisions and clauses, some of which are discussed later in this section. In addition to LPAs, **side letters** may also be negotiated. Side letters are side agreements created between the GP and a certain number of LPs that exist *outside* the LPA. Some examples of clauses or details that may be included in a side letter include the following:

- Potential additional reporting due to an LP's unique circumstances, such as regulatory or tax requirements
- First right of refusal and other similar clauses to outline potential treatment (regarding fees, co-investment rights, secondary sales, and, potentially, other matters) in comparison to other LPs
- Notice requirements in the event of litigation, insolvency, and related matters
- Most favored nation clauses, such as agreeing that if similar LPs pay lower fees, they will be offered to the LP

Certain structures are commonly adopted for specific alternative investments. For example, infrastructure investors frequently enter into public–private partnerships, which are agreements between the public sector and the private sector to finance, build, and operate public infrastructure, such as hospitals and toll roads. In real estate fund investing, investors may be classified as unitholders, and joint ventures are a partnership structure common in real estate direct investing.

KNOWLEDGE CHECK

Fill in the blank: Investments in limited partnerships are ___________ regulated than offerings to the general public.

Solution:

Investments in limited partnerships are **less** regulated than offerings to the general public.

3.2 Compensation Structures

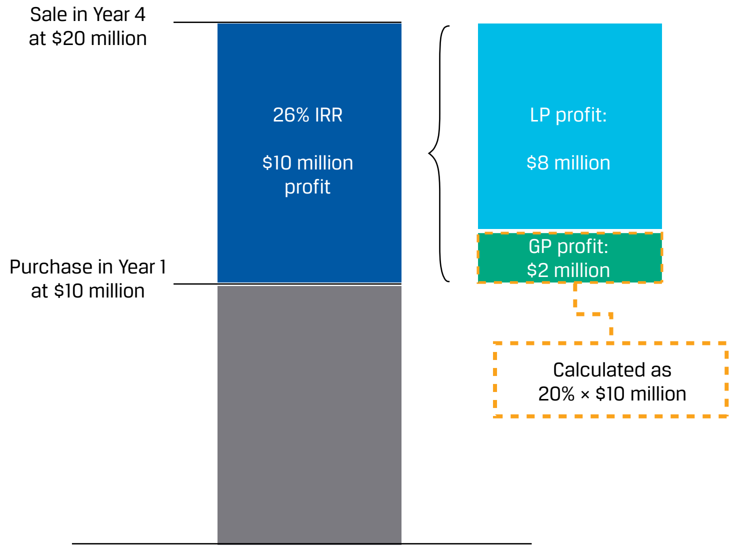
Funds are generally structured with a **management fee** typically ranging from 1% to 2% of assets under management (e.g., for hedge funds) or **committed capital** (e.g., for private equity funds), which is how much money in total that LPs have committed to the fund's future investments. On top of the management fee, a **performance fee** (also referred to as an *incentive fee* or *carried interest*) is applied based on excess returns.

Private equity funds raise committed capital and draw down on those commitments, generally over three to five years, when they have a specific investment to make. Note that the management fee is typically based on committed capital, *not* invested capital; the committed-capital basis for management fees is an important distinction from hedge funds, whose management fees are based on assets under management (AUM). Having committed capital as the basis for management fee calculations reduces the incentive for GPs to deploy the committed capital as quickly as possible to grow their fee base. This allows the GPs to be selective about deploying capital into investment opportunities.

The partnership agreement usually specifies that the performance fee is earned only after the fund achieves a return known as a **hurdle rate**. The hurdle rate is a *minimum* rate of return, typically 8%, that the GP must exceed in order to earn the performance fee. GPs typically receive 20% of the total profit of the private equity fund net of any *hard hurdle rate*, in which case the GP earns fees on annual returns in excess of the hurdle rate, or net of the *soft hurdle rate*, in which case the fee is calculated on the entire annual gross return as long as the set hurdle is exceeded. Hurdle rates are less common for hedge funds but do appear from time to time.

Performance fees are designed to reward GPs for enhanced performance and to motivate investment professionals to work hard and stay involved for years to come. Exhibit 5 illustrates a simple example of how performance fees are calculated using a typical "2 and 20" compensation structure, where management fees are calculated on 2% of AUM or committed capital and performance fees are calculated on 20% of profits. An investment that was purchased in Year 1 for \$10 million and sold in Year 4 for \$20 million represents a return of 26% on an internal rate of return (IRR) basis. The LPs receive 80% of profits (80% × \$10 million profit = \$8 million), and the GP receives 20% of profits (20% × \$10 million profit = \$2 million). Note that this example assumes the hurdle rate has been exceeded and a catch-up clause exists in the partnership agreement; these concepts will be discussed in the next section.

Exhibit 5 Simple Performance Fee Calculation



In addition to both management and performance fees charged to investors, leveraged buyout firms in private equity may charge consulting fees and monitoring fees to the underlying companies.

As for hedge fund managers, they generally accrue an incentive fee on their quarterly performance, but they typically crystalize (realize) their incentive fee once annually. If they make money in the first quarter but then lose the same amount in the second, the investor recoups the Q1 incentive fee and pays only at the end of the year an incentive fee based on total annual returns net of management fees and other expenses.

KNOWLEDGE CHECK

Fill in the blank: A ____________ fee is typically added to a management fee for a fund governed by a limited partnership agreement.

Solution:

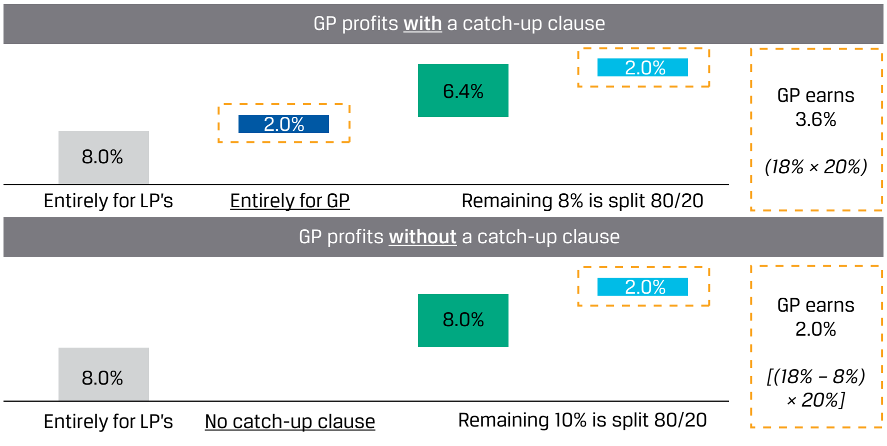
A <u>performance (incentive or carried interest)</u> fee is typically added to a management fee for a fund governed by a limited partnership agreement.

3.3 Common Investment Clauses, Provisions, and Contingencies

For most alternative investment funds, particularly hedge funds and private equity funds, the GP does not earn a performance fee until the LPs have received their initial investment *and* the total return generated on the investment has exceeded a specified hurdle rate. In our example, the LPs receive 80% of the total profit of the fund plus the amount of their initial investment.

A **catch-up clause** may be included in the partnership agreement. Essentially, for a GP who earns a 20% performance fee, a catch-up clause allows the GP to receive 100% of the distributions above the hurdle rate *until* he receives 20% of the profits generated, and then every excess dollar is split 80/20 between the LPs and GP. To illustrate, assume that the GP has earned an 18% IRR on an investment, the hurdle rate is 8%, and the partnership agreement includes a catch-up clause. In this case, the LPs would receive the entirety of the first 8% profit, the GP would receive the entirety of the next 2% profit—because 2% out of 10% amounts to 20% of the profits accounted for so far, as the catch-up clause stipulates—and the remaining 8% would be split 80/20 between the LPs and the GP. Effectively, the LPs earned 14.4% (18% × 80%) and the GP earned 3.6% (18% × 20%). Imagine the same scenario, absent a catch-up clause: The LPs would still receive the entirety of the first 8% profit; however, the remaining 10% would be split 80/20 between the LPs and GP, reducing the GP's return to 2.0% [(18% – 8%) × 20%]. These calculations are presented graphically in Exhibit 6.

Exhibit 6 Simple Catch-Up Clause Illustration



In hedge funds, fee calculations also take into account a **high-water mark**, which reflects the highest value used to calculate an incentive fee. A high-water mark is the *highest* value of the fund investment ever achieved at a performance fee crystallization date, net of fees, by the individual LP. A high-water mark clause states that a hedge fund manager must recuperate declines in value from the high-water mark before performance fees can be charged on newly generated profits. The use of high-water marks protects clients from paying twice for the same performance.

With all alternative investments, investor high-water marks generally carry over into new calendar years, although with hedge funds, an investor will no longer be able to claw back any incentive fees paid for a given calendar year if portfolio losses are subsequently incurred in a later calendar year. Given the generally more illiquid and longer-term nature of their holdings, private equity and real estate investments are more likely to contain clawback clauses for the entire life of the portfolio.

Given that different clients invest at different times, it is possible that not all clients will be at their respective high-water marks at the same time; a client who invests on a dip will enjoy the fund's recovery and pay an incentive fee, whereas a client who invested at the top will need to earn back what was lost before being obliged to pay the incentive fee.

Needless to say, individual capital account fund accounting can become quite complicated when tracking such investment timing differences. Even careful administrators make accounting mistakes when handling high-water mark and incentive fee issues. The astute alternative asset investor will always double-check the incentive fee calculations provided to them. Best-practice accounting for hedge funds is far more complex than the simple issuance of shares (carried at a single net asset value, or NAV), which is more common in the mutual fund industry.

KNOWLEDGE CHECK

The minimum rate of return that a GP must exceed in order to earn an incentive or performance fee is called the:

A high-water mark.

B hurdle rate.

C performance threshold.

Solution:

B is correct. A high-water mark is the highest value used to calculate an incentive fee. "Performance threshold" is not a term that is generally used in the industry.

In alternative investments, a **waterfall** represents the distribution method that defines the order in which allocations are made to LPs and GPs. GPs usually receive a disproportionately larger share of the total profits relative to their initial investment, which incentivizes them to maximize profitability. There are two types of waterfalls: *deal-by-deal* (or *American*) waterfalls and *whole-of-fund* (or *European*) waterfalls. Deal-by-deal waterfalls are more advantageous to the GP because performance fees are collected on a per-deal basis, allowing the GP to get paid before LPs receive both their initial investment *and* their preferred rate of return (i.e., the hurdle rate) on the entire fund. In whole-of-fund waterfalls, all distributions go to the LPs as deals are exited and the GP does not participate in any profits until the LPs receive their initial investment and the hurdle rate has been met. In contrast to deal-by-deal waterfalls, whole-of-fund waterfalls occur at the aggregate fund level (i.e., after *all* investments have been exited) and are more advantageous to the LPs. Exhibits 7 and 8 illustrate deal-by-deal waterfalls and whole-of-fund waterfalls, respectively.

Exhibit 7 Deal-by-Deal (American) Waterfall Example

Investment No.	Year		Amount ($mm)		Profit		
	Invested	Sold	Invested	Sold	$mm	%	GP at 20%
1	1	4	$10	$20	$10	26.0%	$2
2	2	5	$20	$35	$15	20.5%	$3
3	2	7	$40	$80	$40	14.9%	$8
4	3	7	$20	$20	-	-	-
5	3	8	$35	$25	($10)	neg	($2)
6	4	9	$25	$20	($5)	neg	($1)
7	5	9	$30	-	($30)	neg	($6)
8	5	10	$20	-	($20)	neg	($4)
Total	1	10	$200	$200	-	-	-

Exhibit 8 Whole-of-Fund (European) Waterfall Example

Investment No.	Year Invested	Year Sold	Amount ($mm) Invested	Amount ($mm) Sold	Profit $mm	Profit %	Profit GP at 20%
1	1	4	$10	$20	$10	26.0%	-
2	2	5	$20	$35	$15	20.5%	-
3	2	7	$40	$80	$40	14.9%	-
4	3	7	$20	$20	-	-	-
5	3	8	$35	$25	($10)	neg	-
6	4	9	$25	$20	($5)	neg	-
7	5	9	$30	-	($30)	neg	-
8	5	10	$20	-	($20)	neg	-
Total	1	10	$200	$200	-	-	-

A **clawback** provision reflects the right of LPs to reclaim part of the GP's performance fee. Along either waterfall path, if a GP ever accrues (or actually pays itself) an incentive fee on gains that are not yet fully realized and then subsequently gives back these gains, an investor is typically able to claw back prior incentive fee accruals and payments. Clawback provisions are usually activated when a GP exits successful deals early on but incurs losses on deals later in the fund's life.

PRACTICE QUESTIONS FOR LOS C

Question 1

An alternative investment fund's hurdle rate is a:

A rate unrelated to a catch-up clause.

B tool to protect clients from paying twice for the same performance.

C minimum rate of return the GP must exceed in order to earn a performance fee.

Solution:

C is correct. An alternative investment fund's hurdle rate is a minimum rate of return the GP must exceed in order to earn a performance fee.

A is incorrect because if a catch-up clause is included in the partnership agreement, the catch-up clause permits distributions in relation to the hurdle rate.

B is incorrect because it is a high-water mark (not a hurdle rate) that protects clients from paying twice for the same performance.

Question 2

An investor in a private equity fund is concerned that the general partner can receive incentive fees in excess of the agreed-on incentive fees by making distributions over time based on profits earned rather than making distributions only at exit from investments of the fund. Which of the following is most likely to protect the investor from the general partner receiving excess fees? (2020 Q32)

A A high hurdle rate

B A clawback provision

C A lower capital commitment

Solution:

B is correct. A clawback provision requires the general partner in a private equity fund to return any funds distributed (to the general partner) as incentive fees until the limited partners have received their initial investments *and* the contracted portion of the total profits. A high hurdle rate will result in distributions occurring only after the fund achieves a specified return. A high hurdle rate decreases the likelihood of, but does not prevent, excess distributions. Management fees, not incentive fees, are based on committed capital.

Question 3

Until the committed capital is fully drawn down and invested, the management fee for a private equity fund is based on: (2022 Q33)

A invested capital.

B committed capital.

C assets under management.

Solution:

B is correct. Until the committed capital is fully drawn down and invested, the management fee for a private equity fund is based on committed capital, not invested capital.

Question 4

The distribution method by which profits generated by a fund are allocated between LPs and the GP is called:

A a waterfall.

B an 80/20 split.

C a fair division.

Solution:

A is correct. Although profits are typically split 80/20 between LPs and the GP, the distribution method of profits is not called an "80/20 split." "Fair division" is not a real term that exists in the industry.

Question 5

Fill in the blanks with the correct words: An American waterfall distributes performance fees on a(n) ___________ basis and is more advantageous to the ___________.

A deal-by-deal; LPs

B aggregate fund; LPs

C deal-by-deal; GP

Solution:

C is correct. American waterfalls, also known as deal-by-deal waterfalls, pay performance fees after every deal is completed and are more advantageous to the GP because they get paid sooner (compared with European, or whole-of-fund, waterfalls).

4 HEDGE FUNDS

d explain investment characteristics of hedge funds

4.1 Characteristics of Hedge Funds

In 1949, Alfred Winslow Jones, a sociologist investigating fundamental and technical research in order to forecast the stock market for *Fortune* magazine, set up an investment fund with himself as GP. The fund followed three tenets: (1) Always maintain short positions, (2) always use leverage, and (3) charge only an incentive fee of 20% of profits with no fixed fees. Jones called his portfolio a "hedged" fund (eventually shortened to "hedge" fund) because his short positions offset his long positions in the stock market. Theoretically, the overall portfolio was hedged against major market moves.

Although Jones's original three tenets still have some relevance to today's hedge fund industry, not all hedge funds restrict themselves to equities. Many trade sovereign and corporate debt, commodities, futures contracts, options, and other derivatives. Even real estate investments are found in hedge fund structures. Nor do all hedge funds maintain short positions or use leverage. Instead, many simply exploit niche areas of expertise in a sophisticated manner; hedging and leverage may or may not be part of it.

Fee structures have evolved to include management fees in addition to Jones's incentive fee. The performance fees we discussed previously may vary from his simple 20% while including hurdle rates and other mechanisms.

Investors in modern hedge funds are subject to extended holding periods (known as **lockup periods**) and subsequent **notice periods** before an investment redemption is possible. Such mandatory periods combine with a lighter regulatory compliance burden (in certain areas) than mutual funds face, for instance, allowing hedge funds more flexibility. Most hedge funds do not need to guarantee liquidity on one day's redemption notice as the large majority of mutual funds do. Instead, they may avail themselves of less liquid and unnoticed opportunities, the true valuation of which may at times be opaque.

Hedge funds are generally deemed riskier from an oversight (fraud risk) point of view, but some hedge funds take less absolute market risk in their portfolio construction than registered products available to retail investors might take. The distinction between regulatory risk and illiquidity risk when compared with true market risk is often confused.

Overall, the term "hedge fund" today can be seen more as a reference to a certain "vehicle type"—a private limited partnership or offshore fund open only to high-net-worth and institutional investors from a regulatory perspective—that may contain different types of investments and where incentive fees are allowed to be charged. This situation remains quite distinct from the mutual fund world, where incentive fees are typically not allowed and investment and liquidity restrictions are more tightly defined. Although in recent years, there has been some "blurring of the lines" with the introduction of so-called liquid alternative mutual funds and what are known as UCIT funds in Europe (many of which offer more "hedged" styles and approaches than traditional daily-liquidity products), such offerings are not to be confused with true limited-access and limited-liquidity hedge funds.

A contemporary hedge fund generally has the following characteristics:

- It is a creatively managed portfolio of investments involved in one or more asset classes (equities, credit, fixed income, commodities, futures, foreign exchange, and sometimes even hard assets, such as real estate), sometimes trading in different geographic regions, that is often leveraged, generally takes both long and short positions (when possible), and quite often uses derivatives to express a view or establish a hedge.

- Its goal is to generate high returns, either in an absolute sense or on a risk-adjusted basis relative to its portfolio-level volatility. It may be hard to gauge a hedge fund's performance relative to a traditional index benchmark (after all, low correlation with traditional asset investing is frequently a selling point), and it generally enjoys light investment restrictions that are detailed in each hedge fund's private placement offering memorandum.
- It is set up as a private investment partnership or offshore fund that under certain legal restrictions (which vary by jurisdiction) make the offering open to a limited number of investors willing and able to make a substantive initial investment.

As noted previously, restrictions on **redemptions** are typically imposed. Investors may be required to keep their money in the hedge fund for a minimum period (referred to as a lockup period) before they are allowed to make withdrawals or redeem shares. Investors may be required to give notice of their intent to redeem; the notice period is typically 30–90 days. To redeem shares, investors may be charged a fee, typically payable to the fund itself (rather than the manager) so as not to disadvantage remaining investors in the fund, particularly in circumstances where the redemption takes place during the lockup period. This characteristic is called a *soft lockup*, and it offers a path (albeit an expensive one) to redeem early. Again, such restrictions allow hedge fund managers more flexibility in portfolio construction, giving them leeway to invest in situations in which time may be needed to generate an expected return and that, therefore, are not typically suitable for a mutual fund that offers daily liquidity.

Funds of hedge funds are funds that hold a portfolio of hedge funds. They create a diversified portfolio of hedge funds and invite participation from smaller investors or those who do not have the resources, time, or expertise to choose among hedge fund managers. Also, the managers of funds of hedge funds (commonly shortened to "funds of funds") are required to have some expertise in conducting due diligence on hedge funds and may be able to negotiate better redemption or fee terms than individual investors can. Funds of funds may diversify across fund strategies, investment regions, and management styles.

Funds of funds may charge an additional 1% management fee and an additional 10% incentive fee (1 and 10) on top of the fees charged by the underlying hedge funds held in the fund of funds. At both the hedge fund level and the fund-of-funds level, the incentive fee is typically calculated on profits net of management fees.

Fee structures are important to scrutinize here. Fee layering can sorely reduce the initial gross return of an investment in a hedge fund or fund of funds for the end investor. Each hedge fund into which a fund of funds invests is structured to receive a management fee plus an incentive fee. The result may be that the investor is paying fees more than once for the management of the same assets. Finally, liquidity can be of additional concern for investors in funds of funds in times of crisis, when fund redemptions can hurt performance.

KNOWLEDGE CHECK

Fill in the blank: The goal of hedge fund redemption restrictions is typically to ___________ manager flexibility.

Solution:

The goal of hedge fund redemption restrictions is typically to <u>increase</u> hedge fund manager flexibility.

KNOWLEDGE CHECK

Define funds of hedge funds, and identify their primary purpose.

Solution:

Funds of hedge funds are funds that hold a portfolio of hedge funds. Their primary purpose is to add value in manager selection and due diligence and create a diversified portfolio of hedge funds accessible to smaller investors.

Few investors were worried about high fees as hedge funds were achieving average net-of-fee returns of 10%–20% in the 1990s and early 2000s, but as returns migrated toward 6%–9% and the drag on returns from fee structures became much more relevant on a percentage basis, institutional investors challenged these high fees. In late 2013, the largest pension plan in the United States, CalPERS, announced it would curtail its hedge fund allocations partly for this reason.

According to Eurekahedge, by 2018 the average hedge fund fee level had compressed to 1.3% in management fees and 15.5% in incentive fees—considerably less than the 2 and 20 structure envisions but still quite high relative to traditional mutual funds. Fund-of-funds fee levels similarly have compressed; funds of funds can be found charging, for example, either a 1% flat management fee or a 50 bp management fee and only a 5% incentive fee.

Realization of the value-destruction potential of fee layers conspired with increasingly compressed hedge fund returns (for a variety of reasons) to leave funds of funds far less popular in 2020 than they were between 2000 and 2005. Still, a fund of funds may offer compensating advantages, such as access for smaller investors, diversified hedge fund portfolios, better redemption terms, and due diligence expertise.

The overall popularity of hedge funds is illustrated by AUM and net asset flows. Hedge Fund Research, Inc. (HFR, or HFRI in named indexes), reported that AUM grew from approximately $39 billion in 1990 to $491 billion in 2000 and to $3.32 trillion as of December 2019—a clear decade-spanning boom in investment, even if growth has slowed somewhat in the last five years.

Exhibit 9 compares the returns, risk, and performance measures for the HFRI Fund of Funds Composite Index, the MSCI ACWI Index, and the Bloomberg Barclays Global Aggregate Index. The HFRI Fund of Funds Composite Index is an equally weighted performance index of funds of hedge funds included in the HFR Database. Hedge fund indexes suffer from issues related to self-reporting and **survivorship bias**. Survivorship bias relates to the inclusion of only current investment funds in a database. As such, the returns of funds that are no longer available in the marketplace (have been liquidated) are excluded from the database. However, the HFRI Fund of Funds Composite Index reflects the actual performance of portfolios of hedge funds. The measures shown here may reflect a lower reported return because of the added layer of fees, but they likely represent a fairer, more conservative, and more accurate estimate of average hedge fund performance than HFR's composite index of individual funds. The returns are likely just mildly biased toward equity long–short funds since these are always a substantial portion of funds of funds' allocation mix.

As shown in Exhibit 9, over the 25-year period between 1990 and 2014, hedge funds enjoyed higher returns than either stocks or bonds and a standard deviation almost identical to that of bonds. On the basis of the Sortino ratio, hedge funds do not appear as attractive as bonds if returns are adjusted for downside deviation, as reflected in the relative Sortino measures shown in Exhibit 9. The worst **drawdown**, or period of largest cumulative negative returns for hedge funds and global equities, began in 2007 (when each peaked) and ended in 2009.

Notwithstanding this period of stress, when accounting for hedge fund returns' modest overall correlation with global stock returns (0.56) and negligible correlation with global bond returns (0.07) over this 25-year period, hedge funds have certainly offered added value to institutional investors as a portfolio diversification agent.

Exhibit 9 Historical Risk–Return Characteristics of Hedge Funds and Other Investments, 1990–2014

	FoF	Global Stocks	Global Bonds
Annualized return	7.2%	6.9%	6.3%
Annualized volatility	6.0%	16.5%	5.8%
Sharpe ratio	0.63	0.21	0.49
Sortino ratio	0.74	0.43	1.09
Percentage of positive months	69.3%	61.3%	62.7%
Best month	6.8%	11.9%	6.2%
Worst month	−7.5%	−19.8%	−3.8%
Worst drawdown	−22.2%	−54.6%	−10.1%
FoF correlation (avg. monthly)		0.56%	0.07%

Sources: Fund-of-funds (FoF) data are from the HFRI Fund of Funds Composite Index; global stock data are from the MSCI ACWI Index; global bond data are from the Bloomberg Barclays Global Aggregate Index.

Notably, as shown in Exhibit 10, for the subsequent five-year period between 2015 and 2019, the absolute return of funds of funds relative in particular to global equities has declined, while their performance correlation with equity markets actually increased. This trend has made hedge fund allocations arguably less useful and also somewhat less popular. But some allocators have continued to find value to maintain or actually increase their allocation to a mix of hedge funds as a bond market substitute in their overall portfolio building.

Exhibit 10 Historical Risk-Return Characteristics of Hedge Funds and Other Investments, 2015–2019

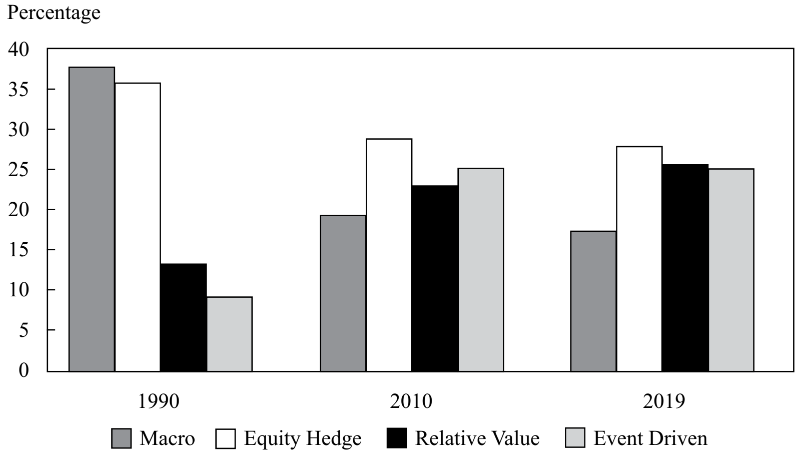
	FoF	Global Stocks	Global Bonds
Annualized return	2.48%	10.41%	2.42%
Annualized volatility	1.08%	11.42%	1.33%
FoF Correlation (avg. monthly)		0.86%	0.03%

Sources: Fund-of-funds (FoF) data are from the HFRI Fund of Funds Composite Index; global stock data are from the MSCI ACWI Index; global bond data are from the Bloomberg Barclays Global Aggregate Index (USD Unhedged). The risk-free rate used in the Sharpe ratio calculation is from the average of one-month T-bill rates over the period.

4.2 Hedge Fund Strategies

Hedge funds are typically classified by strategy, although categorizations vary. Many classifying organizations focus on the most common strategies, but others classify on the basis of different criteria, such as the underlying assets in which the funds are invested. Classifications change over time as new strategies, often based on new products and market opportunities, are introduced. Classifying hedge funds is important so that investors can review aggregate performance data, select strategies with which to build a portfolio of funds, and select or construct appropriate performance benchmarks. HFR classifies hedge funds under four broad strategy categories: event-driven, relative value, macro, and equity hedge. As of March 2019, hedge fund assets stood at approximately \$3.18 trillion. Exhibit 11 shows the approximate percentage of hedge fund AUM by strategy, according to HFR, for 1990, 2010, and 2019. Based on this AUM measure, event-driven and relative value strategies have grown in popularity during the last 30 years, while macro and equity hedge funds have declined.

Exhibit 11 Percentage of AUM by Strategy



Source: HFR data, as of March 2019.

4.2.1 *Equity Hedge Strategies*

Equity hedge strategies can be thought of as the original hedge fund category. They are focused on public equity markets and take long and short positions in equity and equity derivative securities. Most equity hedge strategies use a "bottom-up" security-specific approach—company-level analysis, followed by overall industry analysis, followed by overall market analysis—with relatively balanced long and short exposures. A "top-down" approach, in contrast, entails global macro analysis, followed by sector/regional analysis, followed by individual company analysis or any market-timing approach. Some strategies use individual equities only, whereas others use index ETF securities for portfolio-balancing purposes. Examples of equity hedge strategies include the following:

- **Market neutral.** These strategies use quantitative (technical) and fundamental analysis to identify under- and overvalued equity securities. The hedge fund takes long positions in securities it has identified as undervalued and short positions in securities it has identified as overvalued. The hedge fund tries to maintain a net position that is neutral with respect to market risk and other risk factors (size, industry, momentum, value, etc.). Ideally, the portfolio has an overall beta of approximately zero. The intent is to profit from the movements of individual securities while remaining largely indifferent to market risk. To achieve a meaningful return, market-neutral portfolios may require the application of leverage. They are generally stable, low-return portfolios but are exposed to occasional spurts of volatility, at which point sudden leverage reduction may be required.

- **Fundamental L/S growth.** These strategies use fundamental analysis to identify companies expected to exhibit high growth and capital appreciation. The hedge fund takes long positions in these companies while shorting companies with business models that are being disintermediated or are under downward pressure—where continued revenue growth is deemed unlikely. Long–short growth managers tend to end up long biased overall.
- **Fundamental value.** These strategies use fundamental analysis to identify companies that are deemed undervalued and unloved for any number of corporate-performance or sector-driven reasons, but the manager identifies a path to corporate rejuvenation. The hedge fund takes long positions in these companies and sometimes hedges the portfolio by shorting index ETFs or more growth-oriented companies deemed overvalued. Such managers also tend to be long biased and have a positive factor bias to value as well as to small-cap factors since many underappreciated value situations tend to reside in the small-cap sector.
- **Short biased.** These strategies use quantitative (technical) and fundamental analysis to focus the bulk of their portfolio on shorting overvalued equity securities (against limited long-side exposures or none at all). Many of these funds are more forensic in their fundamental analysis, and some turn activist by trying to expose previously unrecognized accounting or business flaws, thereby also improving their own portfolio's chance for profits. The degree to which these funds adjust their short exposure over time varies by manager. Such managers tend to be contrarian by nature, and the funds can be useful additions to larger portfolios in terms of weathering periods of market stress. Short-biased managers, however, have had a difficult time overall posting meaningful long-term returns during the past 30 years of generally positive market conditions.
- **Sector specific.** These strategies exploit manager expertise in a particular sector and use quantitative (technical) and fundamental analysis to identify opportunities. Technology, biotech/life sciences, and financial services are common areas of focus. Given our era of increased democratization in the dissemination of news, added sector-specialist expertise may be a prerequisite for truly differentiated security selection. The more complex a sector (as in biotechnology, with its binary outcomes on drug trials) or the more opaqueness in accounting practices (as with financial or insurance stocks), the more value sector-specific managers bring.

4.2.2 *Event-Driven Strategies*

Event-driven strategies seek to profit from defined catalyst events, typically those that involve changes in corporate structure, such as an acquisition or restructuring. This strategy is considered to be underpinned by bottom-up security-specific analysis, as opposed to top-down analysis. Investments may include long and short positions in common and preferred stocks, debt securities, and options. Further subdivisions of this category by HFR include the following:

- **Merger arbitrage.** Generally, these strategies involve going long (buying) the stock of the company being acquired at a discount to its announced takeover price and going short (selling) the stock of the acquiring company when the merger or acquisition is announced. The manager may expect to profit once the initial deal spread narrows to the closing value of the transaction after it is fully consummated. The spread exists because there is always some uncertainty over whether the deal will actually close in the face of legal and regulatory hurdles. Equally, a sudden change in market conditions could damage the perceived attractiveness of the merger to the acquirer, who could step away. Shorting the acquirer is also a way to express the risk that the acquirer has overpaid for the acquisition or will suffer from an increased debt load following the merger. The primary risk in merger arbitrage is that the announced merger or acquisition fails to take place and the deal spread re-widens to pre-merger levels before the hedge fund manager is able to unwind its position. Since the expected risk and return on a merger arbitrage strategy typically starts off being quite modest, managers regularly use leverage to amplify merger spreads into financially worthwhile total return targets. Unfortunately, this application of leverage also increases the magnitude of losses should real-world circumstances move against the manager's bet.

- **Distressed/restructuring.** These strategies focus on the securities of companies that are either in bankruptcy or perceived to be near bankruptcy. One strategy sees hedge funds simply purchasing fixed-income securities that are trading at a significant discount to par but are still senior enough to be deemed "money good" (backed by enough corporate assets to be fully repayable at par or at least at a significant premium to the available bond purchase price) in a bankruptcy situation. Alternatively, a hedge fund may purchase the so-called fulcrum debt instrument that is expected to convert into new equity in the case of a restructuring or bankruptcy.
- **Special situations.** These strategies focus on opportunities to get involved in the equity of companies that are engaged in restructuring activities other than mergers, acquisitions, or bankruptcy. These activities include security issuance or repurchase, special capital distributions, rescue finance, and asset sales/spin-offs.
- **Activist.** The term "activist" is short for "activist shareholder." Here, managers secure sufficient equity holdings to allow them to influence a company's policies or direction. The hedge fund manager thus tries to create his or her own catalyst, influencing the investment's ultimate destiny by creating a desired corporate outcome. For example, the activist hedge fund may advocate for divestitures, restructuring, capital distributions to shareholders, or changes in management and company strategy that will affect their equity holdings. Such hedge funds are distinct from private equity because they operate primarily in the public equity market.
- These event-driven strategies tend to be long biased. Although merger arbitrage is the least long biased among them, the strategy still tends to suffer when market conditions weaken because the risk that mergers will fail increases.

4.2.3 *Relative Value Strategies*

Relative value funds seek to profit from a pricing discrepancy, an unusual short-term relationship, between related securities. The expectation is that the pricing discrepancy will be resolved over time. Examples of relative value strategies include the following:

- **Convertible bond arbitrage.** This conceptually market-neutral investment strategy seeks to exploit a perceived mispricing between a convertible bond and its component parts—namely, the underlying bond and the embedded stock option—relative to the pricing of a reference equity into which the bond may someday convert. The strategy typically involves buying convertible debt securities and simultaneously selling a certain amount of the same issuer's common stock. Residual bankruptcy risks can be further hedged using equity put options or credit default swap derivative hedges on the credit of the issuer.

- **Fixed income (general).** These strategies focus on the relative value within the fixed-income markets, with an emphasis on sovereign debt and, at times, the relative pricing of investment-grade corporate debt. Strategies may incorporate long–short trades between two different issuers, between corporate and government issuers, between different parts of the same issuer's capital structure, or between different parts of an issuer's yield curve. Currency dynamics and government yield-curve considerations may also come into play.
- **Fixed income (asset backed, mortgage backed, and high yield).** These strategies focus on the relative value of a variety of higher-yielding securities, including asset-backed securities (ABS), mortgage-backed securities (MBS), and high-yield loans and bonds. The hedge fund seeks both to earn an attractive highly secured coupon and to take advantage of relative security mispricings. At times, some of these securities may be further parsed into structured note products with unique return attributes relative to interest rate movements and general credit spread changes. Opaque **mark-to-market** pricing and illiquidity issues are significant considerations. "Mark-to-market" refers to the current expected fair market value for which a given security would likely be available for purchase or sale if traded in current market conditions, but in the case of structured products, absent hard broker/dealer quotes, this is often a "mark-to-model" type of calculation that is less reliable than an actual bid–offer spread seen between active market participants.
- **Volatility.** These strategies typically use options to go long or short market volatility, either in a specific asset class or across asset classes. For example, a short-volatility strategy involves selling options to earn the premiums and benefit from calm markets but can experience significant losses during unexpected periods of market stress. A long-volatility strategy tends to suffer the cost of small premiums in anticipation of larger market moves where positions may benefit from both directionality and the ability to rebalance option-based exposure through a variety of methods.
- **Multi-strategy.** These strategies trade relative value within and across asset classes or instruments. Rather than focusing on one type of trade (e.g., convertible arbitrage), a single basis for a trade (e.g., volatility), or a particular asset class (e.g., fixed income), this strategy instead looks for investment opportunities wherever they might exist, often with different pods of managers executing their own unique market approaches. The goal of a multi-strategy manager is to initially deploy (and later redeploy) capital efficiently and quickly across a variety of strategy areas as market conditions change.

4.2.4 *Macro and CTA Strategies*

Macro hedge funds emphasize a top-down approach to identify economic trends. Trades are made on the basis of expected movements in economic variables. Generally, these funds trade opportunistically in fixed-income, equity, currency, and commodity markets. Macro hedge funds use long and short positions to profit from a view on the overall direction of the market as it is influenced by major economic trends and events. Macro managers were generally very successful during the 1980s and 1990s, but return profiles have weakened over the past two decades as markets have arguably become more closely managed by central bankers. Because macro managers generally enjoy periods of higher volatility compared with lower volatility, the active moves by national authorities, such as central banks, to smooth out economic shocks likely shrink managers' investment sphere.

Managed futures funds are actively managed funds making diversified directional investments primarily in the futures markets on the basis of a variety of technical and fundamental strategies. Managed futures funds are also known as commodity trading advisers (CTAs) because they historically focused on commodity futures. However, CTAs may include investments in a variety of futures, including commodities, equities, fixed income, and foreign exchange. CTAs generally use models that measure trends and momentum over different time horizons. Investing in CTAs can be useful for portfolio diversification, particularly in times of strong trending market conditions and especially during periods of acute market stress when other fundamental strategies may be expected to perform poorly. However, mean-reverting markets, which may cause false momentum breakout signals, can lead to uncomfortable and occasionally extended drawdown periods before strong trends emerge for the CTA. To the extent that many CTAs have migrated to trade more and more financial products (such as stock index futures and bond futures), the reliability of CTA diversification benefits has diminished.

4.3 Hedge Funds and Diversification Benefits

Hedge funds were generally a niche business throughout much of the 1980s and 1990s, supported mostly by high-net-worth investors. Traditional long–short equity managers shared the spotlight with successful macro-oriented managers, bringing to mind George Soros's Quantum Fund or Julian Robertson's Tiger Management, which took advantage of global currency, equity, and interest rate imbalances between countries.

But when hedge funds generally performed well during the dot-com bubble unwinding of 2000–2002—particularly compared with traditional long-only investment products—endowments, foundations, and pensions started to allocate more money to them. Initially investing in funds of funds, after the 2008 period of market stress, they increasingly made direct allocations. They did so in spite of the high fees charged by hedge funds, and this trend stemmed from an overall effort to achieve better diversification and risk mitigation. This latter era of hedge funds can be characterized by a search for absolute and uncorrelated risk-adjusted returns rather than for outsized upside performance. The institutionalization of the hedge fund industry changed its very nature: Most hedge funds failed to keep pace with the positive equity and bond market advances of 2009–2019, but they maintained a place in institutional asset allocations because of their risk-diversification properties.

PRACTICE QUESTIONS FOR LOS D

Question 1

Which approach is *most commonly* used by equity hedge strategies?

A Top down

B Bottom up

C Market timing

Solution:

B is correct. Most equity hedge strategies use a bottom-up strategy.

A is incorrect because most equity hedge strategies use a bottom-up (not top-down) strategy

C is incorrect because most equity hedge strategies use a bottom-up (not market-timing) strategy.

Question 2

An investor may prefer a single hedge fund to a fund of funds if she seeks: (2022 Q8)

A due diligence expertise.

B better redemption terms.

C a less complex fee structure.

Solution:

C is correct. Hedge funds of funds have multi-layered fee structures, whereas the fee structure for a single hedge fund is less complex. Funds of funds presumably have some expertise in conducting due diligence on hedge funds and may be able to negotiate more favorable redemption terms than an individual investor in a single hedge fund could.

Question 3

Hedge funds are similar to private equity funds in that both:

A are typically structured as partnerships.

B assess management fees based on assets under management.

C do not earn an incentive fee until the initial investment is repaid.

Solution:

A is correct. Private equity funds and hedge funds are typically structured as partnerships where investors are limited partners and the fund is the general partner. The management fee for private equity funds is based on committed capital, whereas for hedge funds, the management fees are based on assets under management. For most private equity funds, the general partner does not earn an incentive fee until the limited partners have received their initial investment back.

Question 4

Both event-driven and macro hedge fund strategies use:

A long–short positions.

B a top-down approach.

C long-term market cycles.

Solution:

A is correct. Long–short positions are used by both types of hedge funds to potentially profit from anticipated market or security moves. Event-driven strategies use a bottom-up approach and seek to profit from a catalyst event typically involving a corporate action, such as an acquisition or a restructuring. Macro strategies seek to profit from expected movements in evolving economic variables.

Question 5

Hedge fund losses are *most likely* to be magnified by a:

A margin call.

B lockup period.

C redemption notice period.

Solution:

A is correct. Margin calls can magnify losses. To meet the margin call, the hedge fund manager may be forced to liquidate a losing position in a security, which, depending on the position size, could exert further price pressure on the security, resulting in further losses. Restrictions on redemptions, such as lockup and notice periods, may allow the manager to close positions in a more orderly manner and minimize forced-sale liquidations of losing positions.

Question 6

An equity hedge fund following a fundamental growth strategy uses fundamental analysis to identify companies that are *most likely* to: (2022 Q16)

A be undervalued.

B be either undervalued or overvalued.

C experience high growth and capital appreciation.

Solution:

C is correct. Fundamental growth strategies take long positions in companies identified, using fundamental analysis, to have high growth and capital appreciation. Fundamental value strategies use fundamental analysis to identify undervalued companies. Market-neutral strategies use quantitative and fundamental analysis to identify under- and overvalued companies.

PRIVATE CAPITAL

5

e explain investment characteristics of private capital

5.1 Overview of Private Capital

Private capital is the broad term for funding provided to companies that is not sourced from the public markets, such as proceeds raised from the sale of equities, bonds, and other securities on exchanges, nor from traditional institutional providers, such as a government or bank. When capital is raised from sources other than public markets and traditional institutions and it is provided in the form of an equity investment, it is called private equity. If similarly sourced capital is instead extended to companies through a loan or other form of debt, it is referred to as private debt. Private capital looks at the entire capital structure, comprising private equity and private debt.

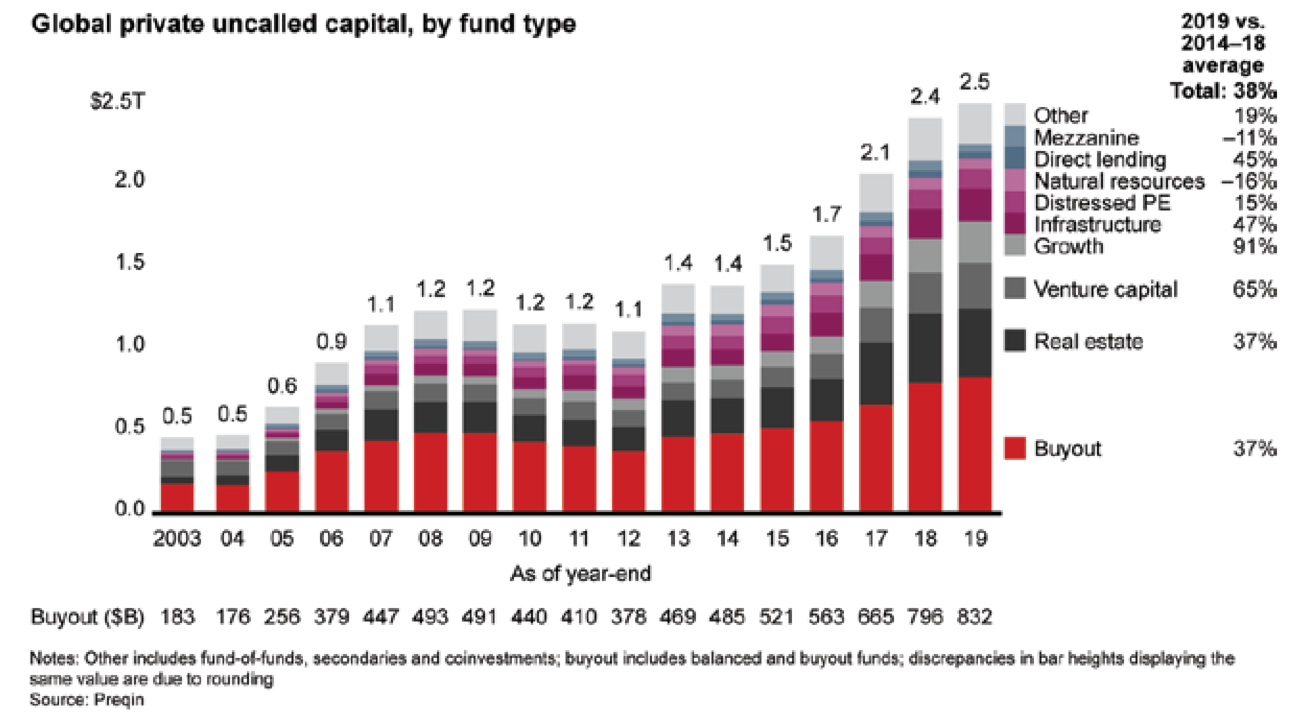
The private capital space largely consists of private investment funds and entities that invest in the equity or debt securities of companies, real estate, or other assets that are privately held. Many private investment firms have private equity and private debt arms; however, these teams typically refrain from investing in the same assets or businesses to avoid overexposure to a single investment and to avoid the conflict of interest that arises from sitting on both sides of the creditors' table. Private investment firms are typically referred to as "private equity firms" even when they have private debt arms. Although private equity is the largest component of private capital, using "private equity" as a generic term could be less accurate and possibly misleading because other forms of private alternative finance have grown considerably in size and popularity.

5.2 Description: Private Equity

Private equity refers to investment in privately owned companies or in public companies with the intent to take them private. As business conditions and the availability of financing change, the focus of private equity firms may change. A firm may manage many private equity funds, each composed of several investments, and the company the firm invests in is often called a **portfolio company** because it will become part of a private equity fund portfolio. According to one possible categorization, we can name leveraged buyouts, venture capital, and growth capital as the primary private equity strategies.

Private equity activity has grown over time, but it is cyclical. Exhibit 12 shows the portion of committed capital that has not yet been called; the industry term for it is "dry powder." Note that detailed information on private equity activity is not always readily available.

Exhibit 12 Global Private Equity: Uncalled Capital by Fund Type (US$ trillions)



Source: Bain & Company, "Global Private Equity Report 2020" (2020).

5.2.1 *Leveraged Buyouts*

Leveraged buyouts (LBOs) or highly leveraged transactions arise when private equity firms establish buyout funds (or LBO funds) to acquire public companies or established private companies, with a significant percentage of the purchase price financed through debt. The target company's assets typically serve as collateral for the debt, and the target company's cash flows are expected to be sufficient to service the debt. The debt becomes part of the target company's capital structure if the buyout goes through. After the buyout, the target company becomes or remains a privately owned company. LBOs are sometimes referred to as "going-private" transactions because after the acquisition of a publicly traded company, the target company's equity is substantially no longer publicly traded. (When the target company is private, it is not a going-private transaction.) The LBO may also be of a specific type. In **management buyouts** (MBOs), the current management team is involved in the acquisition, and in **management buy-ins** (MBIs), the current management team is being replaced and the acquiring team will be involved in managing the company. LBO managers seek to add value by improving company operations, boosting revenue, and ultimately increasing profits and cash flows. Cash-flow growth, in order of contribution, comes from organic revenue growth, cost reductions and restructuring, acquisitions, and then all other sources. The financial returns in this category, however, depend to a large extent on the use of leverage. If debt financing is unavailable or costly, LBOs are less likely to take place.

5.2.2 *Venture Capital*

Venture capital (VC) entails investing in or providing financing to private companies with high growth potential. Typically these are start-ups or young companies, but venture capital can be provided at a variety of stages, ranging from the inception of an idea for a company to the point when the company is about to launch an IPO (initial public offering) or be acquired. The investment return required varies on the basis of the company's stage of development. Investors in early-stage companies will demand higher expected returns relative to later-stage investors because the earlier the stage of development, the higher the risk.

Venture capitalists, like all private equity managers, are not passive investors. They are actively involved with the companies in which they invest.

VC funds typically invest in companies and receive an equity interest but may also provide financing in the form of debt (commonly, convertible debt).

Formative-stage financing is for a company that is still in the process of being formed. Its steps are as follows:

- **a** *Pre-seed capital*, or *angel investing*, is capital provided at the idea stage. Funds may be used to transform the idea into a business plan and to assess market potential. The amount of financing at this stage is typically small and provided by individuals, often friends and family, rather than by VC funds.
- **b** *Seed-stage financing*, or *seed capital*, generally supports product development and marketing efforts, including market research. This is the first stage at which VC funds usually invest.
- **c** *Early-stage financing* (early-stage VC), or *start-up stage financing*, is provided to companies that are moving toward operation but have not yet started commercial production or sales, both of which early-stage financing may be injected to initiate.

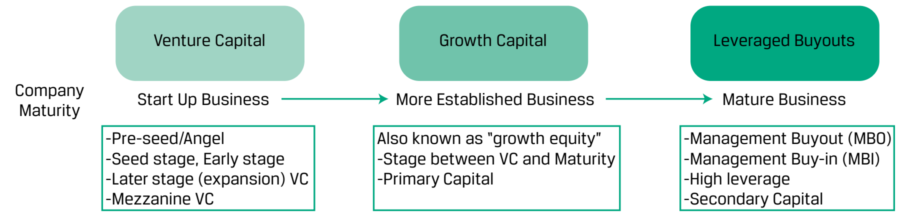
Later-stage financing (expansion VC) is provided after commercial production and sales have begun but before an IPO takes place. Funds may be used to support initial growth, a major expansion (such as a physical plant upgrade), product improvements, or a major marketing campaign.

Mezzanine-stage financing (mezzanine venture capital) prepares a company to go public. It represents the bridge financing needed to fund a private firm until it can complete an IPO or be sold. The term mezzanine-stage financing is used because this financing is provided at the stage between being a private company and being a public company. The focus is on when the financing occurs rather than the financing mechanism itself.

Formative-stage financing is generally carried out by providing ordinary or convertible preferred share sales to the investor or investors, likely the VC fund, while management retains control of the company. Later-stage financing generally involves management selling control of the company to the VC investor; financing is provided through equity and debt, although the fund may also use convertible bonds or convertible preferred shares. The VC fund offers debt financing not for reasons of income generation but, rather, for reasons of recovery and the control of assets in a bankruptcy situation. Simply put, debt financing affords the VC fund more protection than equity does.

To make an investment, a venture capitalist needs to be convinced that the portfolio company's management team is competent and is armed with a solid business plan and strong prospects for growth. Because these investments are not made in mature businesses with years of operational and financial performance history, estimating company valuations on the basis of future prospects is highly uncertain. Accurate estimation is more difficult than in LBO investing, which targets mature, underperforming public companies. Certainty around valuation increases as the portfolio company matures and moves into later-stage financing, but even then, there remains more certainty around LBO investments. Exhibit 13 take us through the growth stages of a company and the types of financing it may receive at each stage:

Exhibit 13 The Private Equity Stage Continuum



Source: Private Equity Primer, "What Is Private Equity" (12 February 2016). www.pe-primer.com/dealsherpapress/2016/2/12/bbcati52gcwdzrl34pts0x2tl783do (accessed June 18, 2020).

5.2.3 *Other Private Equity Strategies*

Among several other specialties, some private equity firms specialize in *growth capital*, also known as growth equity or minority equity investing. Growth capital generally refers to minority equity investments, whereby the firm takes a less-than-controlling interest in more mature companies that are looking for capital to expand or restructure operations, enter new markets, or finance major acquisitions. Many times, minority equity investing is initiated and sought by the management of the investee company, which is interested in realizing earnings from selling a portion of its shares before the company can go public but still seeks to retain control and participation in the success of the company. Although this scenario occurs most commonly with private companies, publicly quoted companies can seek private equity capital through PIPEs (private investments in public equities). Other private equity strategies secure returns by investing in companies in specific industries.

KNOWLEDGE CHECK

Identify two of the four stages of the private equity continuum and the company growth stage each finances.

Solution:

The stages of the entire PE continuum and their associated growth stage financing are as follows:

1 Venture capital focuses on start-up and seed-stage businesses.

2 Growth equity focuses on more established businesses.

3 Recapitalizations focus on more mature businesses, either healthy or distressed.

4 Buyouts/LBOs focuses on mature businesses.

5.2.4 *Exit Strategies*

Private equity firms seek to improve new or underperforming businesses and then exit them at higher valuations, buying and holding companies for an average of five years. The time to exit, however, can range from less than six months to more than 10 years. Before deciding on an exit strategy, private equity managers take into account the dynamics of the industry in which the portfolio company competes, the overall economic cycle, interest rates, and company performance. Managers pursue the following common exit strategies:

- **Trade sale.** This refers to the sale of a company to a strategic buyer, such as a competitor. A trade sale can be conducted by auction or by private negotiation. Benefits include (a) an immediate cash exit for the private equity fund, (b) the potential willingness of strategic buyers to pay more because they anticipate synergies with their own business, (c) fast and simple execution, (d) lower transaction costs than for an IPO, and (e) lower levels of disclosure and higher confidentiality than for an IPO because private equity firms generally deal with only one other party. Disadvantages of trade sales include (a) possible opposition by management (management may wish to avoid being purchased by a competitor for job security reasons), (b) lower attractiveness to employees of the portfolio company than for an IPO (an IPO allows for the monetizing of shares), (c) a limited universe of potential trade buyers, and (d) a potentially lower price for the sale than would be achieved from an IPO.
- **IPO.** In an initial public offering, the portfolio company sells its shares, including some or all of those held by the private equity firm, to public investors. Advantages of an IPO exit include (a) the potential for the highest price, (b) management approval (because management will be retained), (c) publicity for the private equity firm, and (d) future upside potential (because the private equity firm may choose to remain a large shareholder). Disadvantages of an IPO exit include (a) the high transaction costs paid to investment banks and lawyers, (b) long lead times, (c) the risk of stock market volatility (including short-term focus of some investors), (d) onerous disclosure requirements, (e) a potential lockup period (which requires the private equity firm to retain an equity position for a specified period after the IPO), and (f) the fact that IPOs are usually appropriate only for larger companies with attractive growth profiles.
- **Recapitalization.** Recapitalization in the context of private equity describes the steps a firm takes to increase or introduce leverage to its portfolio company and pay itself a dividend out of the new capital structure. A recapitalization is not a true exit strategy, because the private equity firm typically maintains control; however, it does allow the private equity investor to extract money from the company to pay its investors and improve IRR. A recapitalization may be a prelude to a later exit. However, LP investors should be aware that a recapitalization can be used as a method for the GP to manipulate fund IRRs.

- **Secondary sales.** This approach represents a sale of the company to another private equity firm or another group of investors. With the growth of "dry powder" (Exhibit 12), we have seen an increase in the proportion of secondary sales exits.
- **Write-off/liquidation.** A write-off occurs when a transaction has not gone well: The private equity firm revises the value of its investment downward or liquidates the portfolio company before moving on to other projects.

The foregoing exit strategies may be pursued individually or in concert or may be used for a partial exit strategy. For example, private equity funds may sell a portion of a portfolio company to a competitor via a trade sale and then complete a secondary sale to another private equity firm for the remaining portion. Company shares may also be distributed to fund investors, although such a move is unusual.

5.3 Description: Private Debt

Private debt primarily refers to the various forms of debt provided by investors to private entities. In the past decade, the expansion of the private debt market has been largely driven by private lending funds filling the gap between borrowing demand and reduced lending supply from traditional lenders in the face of tightened regulations following the 2008 financial crisis. We can organize the primary methods of private debt investing into four categories: direct lending, mezzanine loans, venture debt, and distressed debt. The broad array of debt strategies offers not only diversification benefits but also exposure to other investment spheres, such as real estate and infrastructure.

5.3.1 *Direct Lending*

Private debt investors get involved in direct lending by providing capital directly to borrowers and subsequently receiving interest, the original principal, and possibly other payments in exchange for their investment. As with typical bank loans, payments are usually received on a fixed schedule, and the loan itself typically is senior and secured and has covenants in place to protect the lender/investor. The loan is provided by a small number of investors to private and sometimes public entities, and it differs from traditional debt instruments, such as bonds, that are issued to many participants and can be publicly traded in the market.

Direct lending primarily involves private debt firms (or private equity firms with private debt arms) establishing funds with money raised from investors looking for higher-yielding debt. Fund managers will then seek opportunities to deploy that capital, such as providing a loan to a mid-market corporation or extending debt to another private equity fund that is seeking financing for acquisitions. In general, private debt funds provide debt, at higher interest rates, to entities that require capital but lack favorable alternatives to traditional bank lenders, who themselves may be uninterested or unable to provide debt to these borrowers. As in private equity, private debt fund managers conduct thorough due diligence before selecting the fund's investments.

In direct lending, many firms may also provide debt in the form of a *leveraged loan*, which is a loan that is itself levered. In other words, private debt firms that invest in leveraged loans will borrow money to finance the debt that the firm then extends to another borrower. By using leverage, a private debt firm can enhance the return on its loan portfolio.

5.3.2 *Mezzanine Debt*

In private debt, mezzanine debt refers to private credit that is subordinated to senior secured debt but is senior to equity in the borrower's capital structure. Mezzanine debt makes a pool of additional capital available to borrowers beyond senior secured debt, and it is often used to finance LBOs, recapitalizations, corporate acquisitions, and similar transactions. Because of its typically junior ranking and the fact that it is usually unsecured, mezzanine debt is riskier than senior secured debt; to compensate investors for this heightened risk profile, investors commonly demand higher interest rates and may require options for equity participation. Mezzanine debt often comes with additional features, such as warrants or conversion rights, which provide equity participation to lenders/investors, meaning they have the option of converting their debt into equity or purchasing the equity of the underlying borrower under certain circumstances.

5.3.3 *Venture Debt*

Venture debt is private debt funding provided to start-up or early-stage companies with venture capital backing that may be generating little or negative cash flow. Entrepreneurs may seek venture debt, which often takes the form of a line of credit or term loan, as a way to obtain additional financing without further diluting shareholder ownership in their business. Venture debt can complement existing equity financing, allowing shareholders to maintain ownership and control over the company for a longer period of time. Similar to mezzanine debt, venture debt may carry additional features that compensate the investor/lender for the increased risk of default or for the fact that start-up and early-stage companies often lack substantial assets that can be pledged as collateral for the debt. One such feature could grant the lender rights to purchase equity in the borrowing company under certain circumstances.

5.3.4 *Distressed Debt*

Involvement in distressed debt typically entails buying the debt of mature companies with financial difficulty. These companies may be in bankruptcy proceedings, have defaulted on debt, or seem likely to default on debt. Some investors identify companies with a temporary cash-flow problem but a good business plan that will help the company survive and, in the end, flourish. These investors buy the company's debt in expectation of both the company and its debt increasing in value. Turnaround investors buy debt and plan to be more active in the management and direction of the company. They seek distressed companies to restructure and revive.

5.3.5 *Other Private Debt Strategies*

Private debt firms may have specialties other than the aforementioned strategies, one of which is investing in or issuing collateralized loan obligations (CLOs), which are leveraged structured vehicles that are collateralized by a portfolio of loans covering a diverse range of tranches, issuers, and industries. A CLO manager extends several loans to corporations—usually to firms involved in LBOs, corporate acquisitions, or similar transactions—pools these loans together, and then divides that pool into various tranches of debt and equity that differ in seniority and security. The CLO manager then sells each tranche to different investors according to their risk profile; the most senior portion of the CLO will be the least risky, and the most junior portion of the CLO (i.e., equity) will be the riskiest.

Another type of debt that could be directly extended to borrowers is **unitranche debt**. Unitranche debt consists of a hybrid or blended loan structure that combines different tranches of secured and unsecured debt into a single loan with a single, blended interest rate. Since unitranche debt is a blend of secured and unsecured debt, the interest rate on this type of loan will generally fall in between the interest rates often demanded on secured and unsecured debt, and the unitranche loan itself will usually be structured between senior and subordinated debt in terms of priority ranking.

Some private debt firms invest in real estate debt or infrastructure debt. *Real estate debt* refers to loans and other forms of debt provided for real estate financing, where a specified real estate asset or property serves as collateral. *Infrastructure debt* encompasses the many forms of debt used to finance the construction, operation, and maintenance of infrastructure assets.

Private debt firms may also provide *specialty loans*, where debt is extended to niche borrowers in specific situations. For example, litigation finance is the practice of a specialist funding company providing debt to clients, usually plaintiffs in litigation, for their legal fees and expenses in exchange for a portion of any case winnings.

KNOWLEDGE CHECK

Which of the following is not considered a strategy in private debt investing?

A Direct lending

B Recapitalization

C Mezzanine debt

Solution:

B is correct. Recapitalization is when a private equity firm increases leverage or introduces it to the company and pays itself a dividend.

KNOWLEDGE CHECK

Which of the following forms of debt are likely to have additional features, such as warrants or conversion rights?

A Mezzanine debt

B Venture debt

C Both A and B

Solution:

C is correct. Both mezzanine and venture debt are likely to have additional features, such as warrants and conversion rights, to compensate debt holders for increased risk, including the risk of default.

5.4 Risk/Return of Private Equity

The higher-return opportunities that private equity funds may provide relative to traditional investments are a function of their ability to invest in private companies, their influence on portfolio companies' management and operations, and their use of leverage. Investing in private equity, including venture capital, is riskier than investing in common stocks. Investors require a higher return for accepting higher risk, including illiquidity and leverage risks.

Exhibit 14 shows the mean annual returns for the Cambridge Associates US Private Equity index, the NASDAQ index, and the S&P 500 Index for a variety of periods ending 31 December 2019. Because private equity returns are based on IRR, however, it is difficult to identify reliable benchmarks for comparing returns to investments.

Some investors may use the public market equivalent (PME) to match the timing of cash flows with the public market. Other investors may use an index of publicly traded private equity firms as a benchmark for private equity returns.

Exhibit 14 Comparison of Mean Annual Returns for US Private Equity and US Stocks

Index	1 Year	5 Years	10 Years	20 Years	25 Years
US private equity	13.95	12.00	14.35	11.08	13.24
NASDAQ	35.23	13.63	14.74	4.03	10.43
S&P 500	31.49	11.70	13.56	6.06	10.22

Source: US Private Equity index and selected benchmark statistics, 31 December 2019, Cambridge Associates.

Published private equity indexes may be an unreliable measure of performance. Measuring historical private equity performance is challenging; as with hedge funds, private equity return indexes rely on self-reporting and are subject to survivorship, backfill, and other biases, which typically lead to an overstatement of returns. Moreover, prior to 2009, in the absence of a liquidity event, private equity firms did not necessarily mark their investments to market. This failure to mark to market leads to an understatement of measures of volatility and correlations with other investments. Thus, data adjustments are required to more reliably measure the benefits of private equity investing. Investors should require a higher return for accepting a higher risk, including illiquidity and leverage risks.

5.5 Risk/Return of Private Debt

Private debt investments may provide higher-yielding opportunities to fixed-income investors seeking increased returns relative to traditional bonds. Private debt funds may generate higher returns by taking opportunistic positions based on market inefficiencies. Indeed, private lending funds stepped into the financing gap left by traditional lenders following the 2008 financial crisis. Investors in private debt could realize higher returns from the illiquidity premium, which is the excess return investors require to compensate for lack of liquidity, and may benefit from increased diversification in their portfolios.

The potential for higher returns is connected to higher levels of risk. Private debt investments vary in risk and return, with senior private debt providing a steadier yield and moderate risk and mezzanine private debt carrying higher growth potential, equity upside and higher risk. As a whole, however, investing in private debt is riskier than investing in traditional bonds. Investors should be aware of these risks, which include illiquidity and heightened default risk when loans are extended to riskier entities or borrowers in riskier situations.

Exhibit 15 shows annualized returns and standard deviations for private debt funds (2004–2016). Based on these figures, private debt funds appear to offer an attractive risk–return trade-off.

Exhibit 15 Historical Annualized Returns and Standard Deviation of Returns for Private Debt Funds, 2004–2016

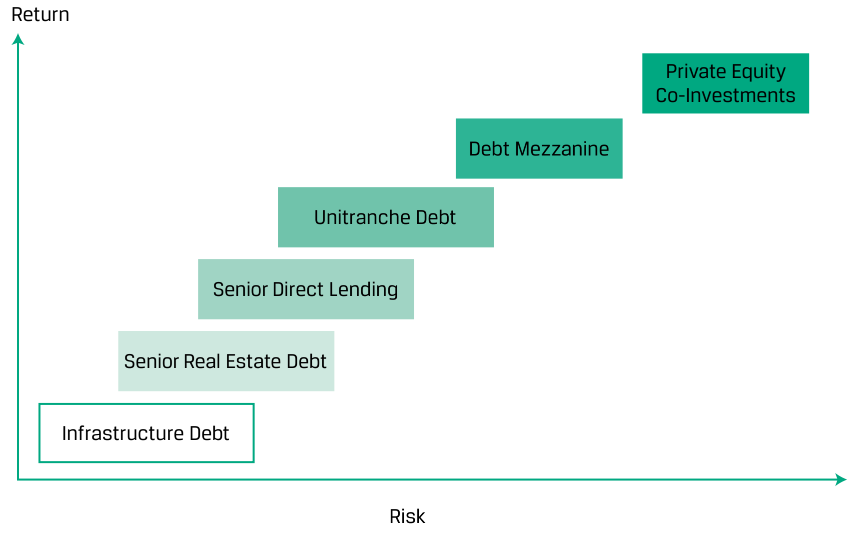
Private Debt Fund	Annualized Return	Standard Deviation
Mezzanine	7.4%	2.5%
Distressed	7.9%	6.7%
Direct lending only:		
All direct lending	9.2%	2.0%
Direct lending (excl. mezzanine)	11.1%	3.4%

Source: S. Munday, W. Hu, T. True, and J. Zhang, "Performance of Private Credit Funds: A First Look," Institute for Private Capital (7 May 2018, pp. 32–3).

Although there is no widely used benchmark for private debt, various indexes may serve investors' purposes as long as they understand the pros and cons of each index. The ICE Bank of America Merrill Lynch Global High Yield Index (HYI) tracks publicly traded, non-investment-grade corporate bonds of large, international corporations; the S&P/LSTA U.S. Leveraged Loan Index (LLI) covers syndicated and over-the-counter traded, non-investment-grade loans for large US corporations; the S&P Net Total Return BDC Index tracks publicly traded business-development companies (BDCs); and the Cliffwater Direct Lending Index (CDLI) is based on quarterly SEC filings covering more than 60 private and public BDCs.

Investments in private equity and private debt vary in terms of risk and return because of various factors, including but not limited to ranking in an entity's capital structure. Exhibit 16 contains a graph that plots private equity and private debt categories by their risk and return levels. (Mirroring the risk–return pathway for traditional equity and debt investing, Exhibit 16 plots private equity and private debt using categories that are broadly similar to how we have defined them. Note the trade-off as investors select between junior and senior debt and between equity and debt.)

Exhibit 16 Private Capital Risk and Return Levels by Category



Source: Based on graph from Leon Sinclair, "The Rise of Private Debt," IHS Markit (7 August 2017).

KNOWLEDGE CHECK

Fill in the blank by choosing from among the following options:

___________ is the type of private debt expected to have the greatest excess return potential.

A Unitranche

B Mezzanine

C Infrastructure

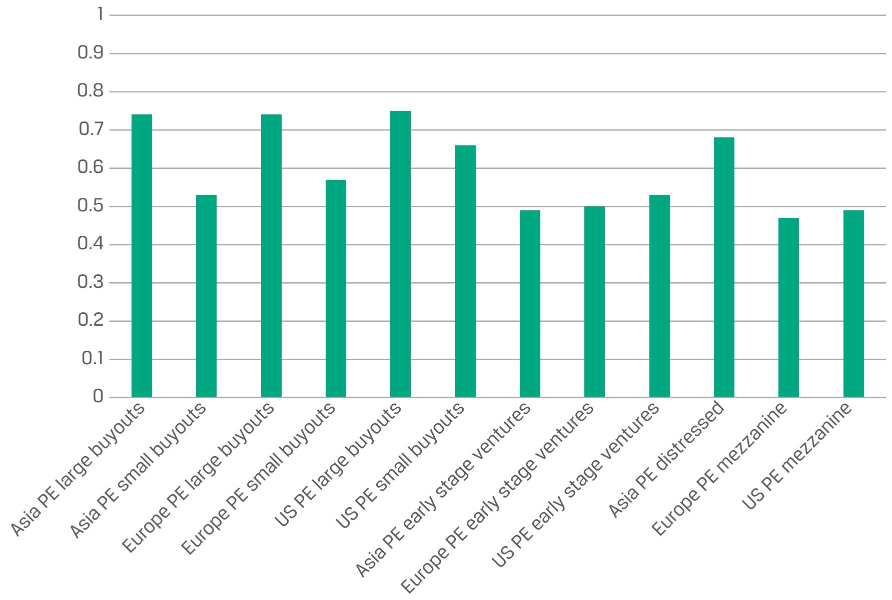
Solution:

B is correct. As a junior form of subordinated debt, mezzanine private debt private debt offers higher growth potential, equity upside, and higher risk, with the comparatively highest returns. Infrastructure debt is senior and poses the lowest risk. Unitranche debt is less risky than subordinated debt but riskier than infrastructure debt and is a blend of secured and unsecured debt; its interest rate generally falls in between the interest rates often demanded on secured and unsecured debt, and the loan itself is usually structured between senior and subordinated debt.

5.6 Diversification Benefits of Investing in Private Capital

Investments in private capital funds can add a moderate diversification benefit to a portfolio of publicly traded stocks and bonds, with correlations with public market indexes varying from 0.47 to 0.75, as shown in Exhibit 17. And if investors identify skillful fund managers, they may benefit from excess returns given the additional leverage, market, and liquidity risks.

Exhibit 17 Private Capital's Average Correlations with Public Market Indexes, 2013–2018



Source: MSCI. www.msci.com/www/blog-posts/did-private-capital-deliver-/01697376382 (accessed 20 July 2020).

Private equity investments may offer vintage diversification because capital is not deployed at a single point in time but invested over several years. Private debt investments, which offer more options than bonds and public forms of traditional fixed income, can also serve diversification efforts. Ultimately, investors should prudently avoid concentration risk, aiming to diversify across different managers, industries, and geographies.

PRACTICE QUESTIONS FOR LOS E

Question 1

A collateralized loan obligation specialist is *most likely* to:

A sell its debt at a single interest rate.

B cater to niche borrowers in specific situations.

C rely on diverse risk profiles to complete deals.

Solution:

C is correct. A CLO manager will extend several loans to corporations (usually to firms involved in LBOs, corporate acquisitions, or other similar types of transactions), pool these loans, and then divide that pool into various tranches of debt and equity that range in seniority and security. The CLO manager will then sell each tranche to different investors according to their risk profiles; the most senior portion of the CLO will be the least risky, and the most junior portion of the CLO (i.e., equity) will be the riskiest.

A is incorrect because with the different CLO tranches having distinct risks varying with their seniority and security, they will be priced over a range of interest rates. In contrast, unitranche debt combines different tranches of secured and unsecured debt into a single loan with a single, blended interest rate.

B is incorrect because debt extended to niche borrowers in specific situations is more commonly offered through specialty loans. For example, in litigation finance, a specialist funding company provides debt to a client to finance the borrower's legal fees and expenses in exchange for a portion of any case winnings.

Question 2

Private capital is:

A accurately described by the generic term "private equity."

B a source of diversification benefits from both debt and equity.

C predisposed to invest in both the debt and equity of a client's firm.

Solution:

B is correct. Investments in private capital funds can add diversity to a portfolio composed of publicly traded stocks and bonds because they have less-than-perfect correlation with those investments. There is also the potential to offer further diversification within the private capital asset class. For example, private equity investments may also offer vintage diversification since capital is not deployed at a single point in time but is invested over several years. Private debt provides investors with the opportunity to diversify the fixed-income portion of their portfolios since private debt investments offer more options than bonds and other public forms of traditional fixed income.

A is incorrect because although private equity is considered by many to be the largest component of private capital, using "private equity" as a generic term could be less accurate and possibly misleading since other private forms of alternative finance have grown considerably in size and popularity.

C is incorrect because although many private investment firms often have private equity and private debt arms, these teams typically won't invest in the same assets or businesses to avoid overexposure to a single investment.

Question 3

The first stage of financing at which a venture capital fund *most likely* invests is the: (2020 Q13)

A seed stage.

B mezzanine stage.

C angel investing stage.

Solution:

A is correct. The seed stage supports market research and product development and is generally the first stage at which venture capital funds invest. The seed stage follows the angel investing stage. In the angel investing stage, funds are typically provided by individuals (often friends or family), rather than a venture capital fund, to assess an idea's potential and to transform the idea into a plan. Mezzanine-stage financing is provided by venture capital funds to prepare the portfolio company for its IPO.

Question 4

A private equity fund desiring to realize an immediate and complete cash exit from a portfolio company is *most likely* to pursue: (2020 Q18)

A an IPO.

B a trade sale.

C a recapitalization.

Solution:

B is correct. Private equity funds can realize an immediate cash exit in a trade sale. Using this strategy, the portfolio company is typically sold to a strategic buyer.

Question 5

Angel investing capital is typically provided in which stage of financing? (2020 Q24)

A Later stage

B Formative stage

C Mezzanine stage

Solution:

B is correct. Formative-stage financing occurs when the company is still in the process of being formed and encompasses several financing steps. Angel investing capital is typically raised in this early stage of financing.

Question 6

Private equity funds are *most likely* to use:

A merger arbitrage strategies.

B leveraged buyouts.

C market-neutral strategies.

Solution:

B is correct. The majority of private equity activity involves leveraged buyouts. Merger arbitrage and market neutral are strategies used by hedge funds.

6 NATURAL RESOURCES

f explain investment characteristics of natural resources

6.1 Overview of Natural Resources

We can define natural resources as a unique asset category comprising commodities and raw land used for farming and timber. Some investors also include real estate assets and infrastructure in this category to form a "real assets" asset class, which can include inflation-protected securities and miscellaneous investments designed to protect against an increase in consumer prices. Others include timberland and farmland in their real estate portfolios. Regardless, the particular investment characteristics of natural resources, the specialized knowledge that their investors require, and their inclusion in more portfolios today than a decade ago call for a separate examination of the sector.

Commodities are broadly familiar to both professional investors and the general public: Crude oil, soybeans, copper, and gold are all commodities. Commodities are considered either “hard” (those that are mined, such as copper, or extracted, such as oil) or “soft” (those that are grown over a period of time, such as livestock, grains, and cash crops, such as coffee). Timberland investment involves ownership of raw land and the harvesting of its trees for lumber, thus generating an income stream and the potential for capital gain; it has formed part of large institutional portfolios for decades. Farmland as an investment is a more recent phenomenon, with only a few dedicated funds involved. With population growth, weather, and water management becoming more topical, however, investors may turn to these sustainable land assets to address such concerns in their portfolio.

An examination of natural resources investment requires an important clarification up front: The aspects discussed here are associated with investing in the physical land and products that come from that land—petroleum, metals, grains—not from the companies that produce those products. Until about 20 years ago, the only commonly available investment vehicles related to this asset class were, indeed, financial instruments, such as stocks and bonds. Nowadays, the wide variety of direct investments via ETFs, limited partnerships, REITS, swaps, and futures opens the door for almost everyone to invest in these assets directly.

This section covers both commodities and land (farmland and timberland) in each subsection, leaving the broader real estate category for the next section.

6.2 Characteristics of Natural Resources

Here, we offer an overview and describe properties that separate these investments from the more commonly known asset classes addressed earlier in this reading.

6.2.1 *Commodities*

Commodities are physical products that can be standardized on quality, location, and delivery for the purpose of investing. Returns on commodity investments are primarily based on changes in price rather than on income from interest, dividends, or rent. In fact, holding commodities (i.e., the physical products) incurs transportation and storage costs. Therefore, trading in physical commodities is primarily limited to a smaller group of entities that are part of the physical supply chain. Most commodity investors do not trade actual physical commodities but, rather, trade commodity derivatives. The underlying asset of the commodity derivative may be a single commodity or an index of commodities.

Because the prices of commodity derivatives are, to a significant extent, a function of the underlying commodity prices, it is important to understand the physical supply chain and general supply–demand dynamics. In fact, the price volatility associated with commodity derivative indexes is highly correlated with the underlying physical goods. This fact should make intuitive sense because the supply chain participants use futures to hedge their forward purchases and sales of the physical commodities. Imagine how uncertain food prices would be if a soybean farmer could not hedge his crop risk. Investors, sometimes referred to as speculators, trade commodity derivatives in search of profit based largely on changes or expected changes in the price of the underlying commodities. Such non-hedging investors include retail and institutional investors, hedge funds, proprietary desks within financial institutions, and trading desks operating in the physical supply chain. (Yes, commodity producers and consumers both hedge *and* speculate on commodity prices.)

Commodity sectors include precious and base (i.e., industrial) metals, energy products, and agricultural products. Exhibit 18 offers examples of each type. The relative importance, amount, and price of individual commodities evolve with society's preferences and needs. For example, the increasing industrialization of China, India, and other emerging markets has driven strong global demand for commodities. Developing markets need increasing amounts of oil, steel, and other materials to support manufacturing, infrastructure development, and the consumption demands of their populations. Emerging technologies, such as advanced cell phones and electric vehicles, create demand for new materials and destroy demand for old resources. The supply and demand for specific commodities evolves over time.

Exhibit 18 Examples of Commodities

Sector	Sample Commodities
Energy	Oil, natural gas, electricity, coal
Base metals	Copper, aluminum, zinc, lead, tin, nickel
Precious metals	Gold, silver, platinum
Agriculture	Grains, livestock, coffee
Other	Carbon credits, freight, forest products

Commodities may be further classified on the basis of physical location and grade or quality, for example. There are many grades and delivery locations for crude oil, and there are many grades and delivery locations for wheat. Commodity derivative contracts, therefore, specify quantity, quality, maturity date, and delivery location.

KNOWLEDGE CHECK

Which of the following can be considered as an investment in the commodities asset class?

A A set of rare antique coins, some of which are made of silver and gold

B A lease on an oil tanker shipping oil between Saudi Arabia and China

C Ownership of a battery factory that uses industrial metals

D A metric tonne of coffee packed in jute bags in a warehouse in Brazil

Solution:

D is correct, assuming the investor owns the commodity directly, not via a derivatives contract. A describes an investment in collectables (whose value has more to do with the rarity of the collectible rather than the actual value of the silver or gold). B describes a financial contract—effectively, a lease or bond. C is also a financial contract that pays on the use of commodities but depends on many other factors (technology, marketing, rent, and employees) for returns.

In order to be transparent, investable, and replicable, commodity indexes typically use the price of the futures contracts rather than the prices of the underlying commodities. As a result, the performance of a commodity index can differ from the performance of the physical commodities. Different commodity indexes are composed of different commodities and weight their component commodities differently. Thus, exposures to specific commodities and commodity sectors vary. However, the low correlation between commodities and other asset classes (e.g., stocks and bonds) means that an investor can achieve improved portfolio diversification regardless of the index she chooses.

Futures are the basis for the vast majority of commodity investment, and each contract is for the future delivery or receipt of a set amount of the commodity in question—for example, 1,000 barrels of oil or 10 metric tonnes of cocoa. This means that the price can be formalized in the following form:

$$\text{Futures price} \approx \text{Spot price}(1 + r) + \text{Storage costs} - \text{Convenience yield},$$

where r is the period's short-term risk-free interest rate. The collateral posted for the contract should also be included if a total return calculation is required. The storage and interest costs together are sometimes referred to as the "cost of carry" or the "carry." We recognize that the buyer of the futures contract has no immediate access to the commodity and, therefore, cannot benefit from it, and so the futures price is adjusted for the loss of convenience, a value that varies. For example, the convenience yield from possessing heating oil in in Finland during the winter is higher than the convenience yield from possessing heating oil in Finland during the summer.

Futures prices may be higher or lower than spot prices depending on the convenience yield. When futures prices are higher than the spot price, the commodity forward curve is upward sloping, and the prices are referred to as being in **contango**. Contango generally occurs when there is little or no convenience yield. When futures prices are lower than the spot price, the commodity forward curve is downward sloping, and the prices are referred to as being in **backwardation**. Backwardation occurs when the convenience yield is high. As a rule of thumb, a contango scenario generally lowers the return of the long-only investor, and a backwardation scenario enhances it.

The pricing of derivatives and the theories of commodity pricing are covered in other readings in the CFA Program curriculum.

6.2.2 *Timberland and Farmland*

Real estate property ownership is represented by a title and may reflect access to air rights, mineral rights, and surface rights in addition to building and land-use rights. For the purposes of this section, we are discussing land owned or leased for the benefit of the returns it generates from crops and timber. Given that these resources consume carbon as part of the plant life cycle, their value comes not just from the harvest but also from the offset to human activity. Water rights are also part of the direct and implied value of these properties; conservation easements may create value by supporting traditions and nature conservation. As interest in investments that adhere to environmental, social, and governance (ESG) considerations grows, arable land may fit the criteria.

Sustained interest in these investments stems from their global nature (everyone eats, everyone requires shelter), the income generated from selling crops, inflation protection from holding land, and the degree of insulation they offer from financial market volatility. US farmland, for example, enjoyed positive returns during the two periods after World War II when US GDP declined significantly (1973-1975 and 2007-2009) and during the three periods after World War I (1915-1920, 1940-1951, and 1967-1981) when the United States experienced higher-than-normal inflation. Timberland has been part of institutional and ultra-high-net-worth portfolios for decades, typically trading in larger units of land. Farmland can be found in much smaller sizes—perhaps tens of or a few hundred acres. Many farms are still family owned, as is 98% of US farmland. One of the main challenges of these investments cited by industry participants is their long market cycle, particularly in new-growth forest and crops that are picked, such as fruit.

Timberland offers an income stream based on the sale of trees, wood, and other timber products and has been not highly correlated with other asset classes. Timberland can be thought of as both a factory and a warehouse. Timber (trees) can be grown and easily stored by simply not harvesting them. This characteristic offers the flexibility of harvesting more trees when timber prices are up and delaying harvests when prices are down. The three primary return drivers are biological growth, changes in spot prices and futures prices of lumber (cut wood), and changes in the price of the underlying land.

Farmland is often perceived to provide a hedge against inflation. Similar to timberland, the returns include an income component related to harvest quantities and agricultural commodity prices. Farmland consists mainly of row crops that are planted and harvested (i.e., more than one round of planting and harvesting can occur in a year depending on the crop and the climate) and permanent crops that grow on trees (e.g., nuts) or vines (e.g., grapes). Unlike timberland, farm products must be harvested when ripe, so there is little flexibility in production. Therefore, commodity futures contracts can be combined with farmland holdings to generate an overall hedged return. Recall that a farm is inherently "long" the crop and, therefore, will sell futures that require delivery at the time of the harvest. Farmland may also be used as pastureland for livestock. Similar to timberland, farmland has three primary return drivers: harvest quantities, commodity prices (e.g., the price of corn), and land price changes.

KNOWLEDGE CHECK

Alexandra is considering buying a tract of farmland for long-term capital appreciation and current income. Which of the following factors play a role in evaluating the attractiveness of the investment?

A The land's rights to preferential water access

B The land's proximity to a large, privately held nature reserve

C The land's soil chemical composition allowing the growth of a variety of crops

D All of the above

Solution:

The answer is D. A is important to ensure that crops can be grown with a reasonable chance of success. B allows for the land to be rented to the stewards of the nature reserve and to lay fallow instead of growing food for human consumption. Nature preserve foundations are interested in the land around them because (1) runoff and pesticide use affect the plants and animals in the preserve, (2) migration of animals in the preserve may incur onto the farmland, and (3) foundations often look to add to their holdings, allowing for a future sale of the land. C allows for the possibility of being able to react to market preferences and more stable productivity and thus more reliable income.

KNOWLEDGE CHECK

Large institutional investors consider timberland investments because:

A The small parcel sizes permit fine tuning of their holdings across geography and wood types.

B The optionality around harvesting gives investors the choice between cutting trees for lumber and current income or letting them grow another year for future gain.

C The short return history allows for many alpha opportunities by knowledgeable active managers.

D Clear-cutting trees and destroying nature is appreciated by ESG investors, which are becoming a larger portion of the investment universe.

Solution:

B is the correct answer. A is incorrect because the parcel sizes are generally large, especially compared with farmland. C is incorrect because the return history is relatively long, not short. D states the opposite reason why ESG investors may be interested in timberland investments—for the opportunity to create "conservation zones."

6.3 Risk/Return of Natural Resources

Although we present commodities, farmland, and timberland in the same section, they do have different return drivers and cycles. Commodities are priced on a second-by-second basis on public exchanges, whereas land generally has an infrequent pricing mechanism and may include estimates as opposed to actual transactions. Keeping these market structure differences in mind will help investors consider their relative benefits and challenges.

6.3.1 *Risk/Return: Commodities*

The arguments for investing in commodities include the potential for returns, portfolio diversification, and inflation protection. Investors may choose commodities if they believe prices will increase in the short or intermediate term. Since commodity prices directly feed into inflation index level calculations, commodities serve as a real hedge against inflation risk even when they yield little or no real return. Commodity futures contracts may offer investors a liquidity premium or other trading opportunities, creating the prospect for a real return greater than zero. Note that commodity investments, especially when combined with leverage, exhibit high volatility (see Exhibit 19), which has led to many well-publicized losses among commodity players.

Exhibit 19 Historical Returns of Commodities, 1990–Q1 2020 (quarterly data)

	Global Stocks	Global Bonds	Commodities
Annualized return	4.8%	5.3%	−1.7%
Annualized standard deviation	16.2%	6.0%	23.8%
Worst calendar year	−43.5%	−5.2%	−46.5%
	(2008)	(1999)	(2008)
Best calendar year	31.6%	19.7%	49.7%
	(2003)	(1995)	(2000)

Sources: Global stocks = MSCI ACWI; global bonds = Bloomberg Barclays Global Aggregate Index; commodities = S&P GSCI Total Return.

Before investors dismiss commodities as a source for returns, let us examine the period before the Global Financial Crisis (GFC) in 2008–2009. If we look at the world economy before central banks began injecting trillions of dollars, euros, yuan, yen, and so on, in liquidity and capital market purchases, we can see that commodities held their own in terms of returns. In fact, commodities well outperformed global stocks. Therefore, asset allocators need to ask whether conditions or sentiment can return to the pre-GFC situation when inflation was a factor and government bonds yielded mid- to high single digits.

Exhibit 20 Historical Returns of Commodities, 1990–Q4 2007 (quarterly data)

	Global Stocks	Global Bonds	Commodities
Annualized return	5.2%	7.0%	6.8%
Annualized standard deviation	17.0%	6.3%	18.3%
Worst calendar year	–20.5%	–5.2%	–35.7%
	(2002)	(1999)	(1998)
Best calendar year	31.6%	19.7%	49.7%
	(2003)	(1995)	(2000)

Sources: Global stocks = MSCI ACWI; global bonds = Bloomberg Barclays Global Aggregate Index; commodities = S&P GSCI Total Return.

Commodity spot prices are a function of supply and demand, the costs of production and storage, the value to users, and global economic conditions. Supplies of physical commodities are determined by production and inventory levels and secondarily by the actions of non-hedging investors. Demand for commodities is determined by the needs of end users and secondarily by the actions of non-hedging investors. Investor actions can both dampen and stimulate commodity price movements, at least in the short term.

Producers cannot alter commodity supply levels quickly because extended lead times are often needed to affect production levels. For example, agricultural output may be altered by planting more crops and changing farming techniques, but at least one growing cycle is required before there is an effect. And for agricultural products, at least one factor that is outside the producer's control—the weather—will significantly affect output. Building the necessary infrastructure for increased oil and mining production may take many years: It is a matter not just of developing the mine itself but also of the necessary transportation and smelting components. For commodities, suppliers' inability to quickly respond to changes in demand levels may result in supply levels that are too low in times of economic growth and too high when the economy slows. And despite advancing technology, the cost of new supply may grow over time. For example, new energy and mineral exploration is frequently more expensive because the easy discoveries tend to be exploited first. If fixed production costs are high, producers are unlikely to produce more than what is needed to meet anticipated demand, and they are unlikely to maintain more than modest levels of inventory, leading to the risk of shortages and price spikes.

Overall demand levels are influenced by global manufacturing dynamics and economic growth. Manufacturing needs can change in a period of months as orders and inventories vary. Investors seek to anticipate these changes by watching inventories closely and monitoring economic conditions, including those relating to government policy and growth forecasts. When demand levels and investors' orders to buy and sell during a given period change quickly, the resulting mismatch of supply and demand may lead to price volatility.

6.3.2 *Risk/Return: Timberland and Farmland*

Turning to land, Exhibit 21 provides a comparison of returns on US timber and farmland. The National Council of Real Estate Investment Fiduciaries (NCREIF) constructs a variety of appraisal-based indexes for property, timber, and farmland. Over the 1990–Q1 2020 period, farmland had the highest annualized return and timber had the highest standard deviation.

Exhibit 21 Historical Returns of US Real Estate Indexes, 1990–Q1 2020, Quarterly Data

	NCREIF Data	
	Timber	**Farmland**
Annualized return	9.1%	11.0%
Annualized standard deviation	7.0%	5.9%
Worst calendar year	−5.2%	2.0%
	(2001)	(2001)
Best calendar year	37.3%	33.9%
	(1992)	(2005)

Although the data in Exhibit 21 make farmland appear to be a very attractive investment, they do not tell the whole story. Liquidity is very low and the risk of negative cash flow high because fixed costs are relatively high (remember that the land must be cared for and crops need fertilizer, seed, and so on) and revenue is highly variable based on the weather. The risks of timber and farmland are similar to those of other real estate investments in raw land, but we highlight weather as a unique and more exogenous risk for these assets compared to traditional commercial and residential real estate properties. Drought and flooding can dramatically decrease the harvest yields for crops and thus the income stream expected by investors. The second primary risk is the international competitive landscape. Although real estate is often considered a local investment, productive land generates commodities that are globally traded and consumed. For example, there have been interruptions in world trade, and growing agricultural competition has resulted in declining grain prices. Therefore, looking ahead, these returns seem unlikely to be repeated in the next 30 years. Timber and farmland investments should consider the international context as a major risk factor.

KNOWLEDGE CHECK

Describe a risk of farmland that distinguishes it from real estate investment in raw land.

Solution:

There are two significant risks differentiating farmland from raw land investment:

1. Weather is a more unique and exogenous risk for farmland, with drought or flooding dramatically decreasing many crop yields and thus the expected income stream.
2. Productive land generates globally traded and consumed commodities. This international competitive landscape can result in interruptions in world trade, growing foreign agricultural competition, and resulting declines in crop prices.

6.4 Diversification Benefits of Natural Resources

Although they often entail higher transaction costs and higher informational hurdles, alternative assets generally offer diversification as a major benefit. Where that diversification comes from needs to be understood in order for managers to better set investor expectations.

6.4.1 *Diversification Benefits: Commodities*

Commodity investing may be attractive to investors not only for the potential profits but also because of the following perceptions: (1) Commodities are effective hedges against inflation, which is to say that commodity prices have historically been positively correlated with inflation, because they are an input to prices, and (2) commodities are effective for portfolio diversification, which is to say that commodity returns have historically had a low correlation with other investment returns. Institutional investors, particularly endowments, foundations, and increasingly corporate and public pension funds and sovereign wealth funds, are allocating more of their portfolios to commodities and commodity derivatives.

The portfolio diversification argument is based on the observation that commodities historically have behaved differently from stocks and bonds during the business cycle. Exhibit 22 shows the quarterly correlation between selected commodity, global equity, and global bond indexes. In the period from 1990 through Q1 2020, commodities exhibited a low correlation with traditional assets; the correlations of commodities with global stocks and global bonds were 0.37 and 0.05, respectively. The correlations of stocks, bonds, and commodities are expected to be positive because each of the assets has some exposure to the global business cycle. Note that the selected commodity index, the S&P GSCI (Goldman Sachs Commodity Index), is heavily weighted toward the energy sector and that each underlying commodity may exhibit unique behavior.

Exhibit 22 Historical Commodity Return Correlations, 1990–Q1 2020 (quarterly data)

	Global Stocks	Global Bonds	Commodities	US CPI
Global stocks	1	0.09	0.37	0.21
Global bonds		1	0.05	−0.03
Commodities			1	0.64

Sources: Global stocks = MSCI ACWI; global bonds = Bloomberg Barclays Global Aggregate Index; commodities = S&P GSCI Total Return.

The argument for commodities as a hedge against inflation is related to the fact that some commodity prices are a component of inflation calculations. Commodities, especially energy and food, affect consumers' cost of living. The positive correlation between quarterly commodity price changes and quarterly changes in the US CPI of 0.64 supports this assertion. In contrast, the quarterly return correlations between the US CPI and global stocks and global bonds are close to zero. The volatility of commodity prices, especially energy and food, is much higher than that of reported consumer inflation. Note that consumer inflation is computed from many products used by consumers, including housing, that change more slowly than commodity prices, and inflation calculations use statistical smoothing techniques and behavioral assumptions. However, even in the recent period of low inflation (or even negative inflation in many countries), investing in commodities outperformed global stocks and bonds on average during the same calendar year if inflation was relatively high (i.e., US CPI greater than 2%; see Exhibit 23).

Exhibit 23 Historical Asset Class Returns When US CPI Was >2% versus <2%, 1990–2019 (quarterly data)

	Global Stocks	Global Bonds	Commodities
Higher inflation	+10.9%	+7.1%	+15.6%
Lower inflation	+0.3%	+2.7%	−20.0%

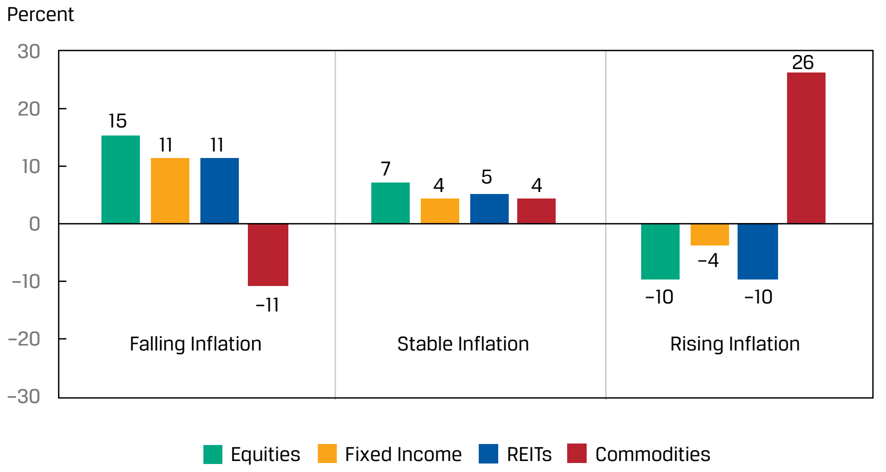
Sources: Global stocks = MSCI ACWI; global bonds = Bloomberg Barclays Global Aggregate Index; commodities = S&P GSCI Total Return.

In addition, Exhibit 23 affirms the sensitivity of commodities to inflation as measured by its relatively high correlation.

Note that changes in commodity prices will coincidentally affect inflation because those changes are included in inflation measures. However, in looking at the weight of food and energy in inflation calculations, they are a relatively small portion. A review of the US CPI calculations indicates an approximate 13% weight of the commodity price impacts. Furthermore, this number is overstated because these price changes as measured by CPI include all the other costs to consumers, such as gasoline taxes, supermarket rent and labor costs, and brand marketing. Therefore, even though there is a direct impact from changes in commodity costs on consumer inflation, it is marginal for the purposes of this discussion.

Exhibit 24 outlines a number of common assets in various regimes of inflation with certain sector callouts that highlight their impact. One example is that mining equities have performed notably better than overall equities in periods of high inflation because of the operational leverage inherent in mining businesses (high fixed costs, low operating costs). This example again highlights the diversification opportunities that can exist between commodities and traditional asset classes.

Exhibit 24



Source: Wellington Management Company LLP

6.4.2 *Diversification Benefits: Timberland and Farmland*

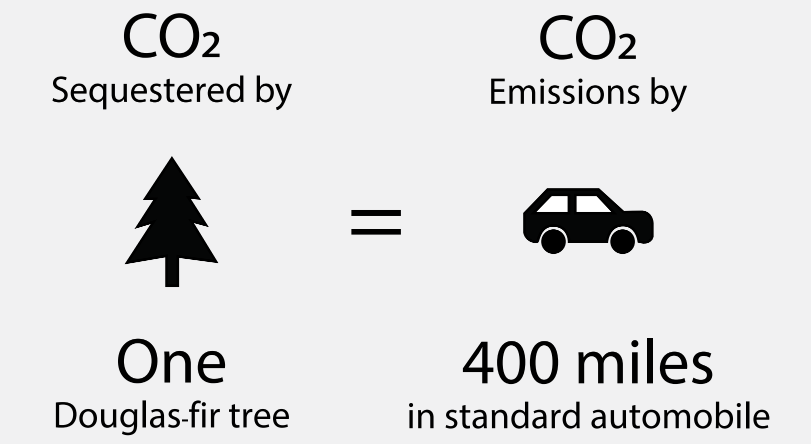
Investors look at diversification in many forms; the numbers tell part of the story, but there are qualitative considerations as well. For example, investors can adhere to ESG principles of responsible and sustainable investing by including timberland and farmland in their portfolios. The following example provides a case study on timberland assets and how investing in timberland can help mitigate climate change.

TIMBERLAND CASE STUDY: CAMPBELL GLOBAL

Investing Responsibly in Timberland Assets: A Climate-Conscious Case Study

Campbell Global (CG) is a global investment manager focused on forest and natural resources investments. Based in Portland, Oregon, with offices in 14 US states and New Zealand, the firm has nearly four decades of experience in sustainable value creation. CG is committed to managing its forests in a manner that promotes the best long-term interests of its clients, while also striving to address economic and ESG considerations.

Figure 1



Source: Campbell Global, LLC

In addition to their economic value, forests serve as vast carbon sinks, with trees removing CO_2 from the atmosphere and using it as carbon storage. In one year, a single Douglas fir tree, a common commercial timber species in the US Pacific Northwest, stores the CO_2 equivalent of driving 400 miles in a standard automobile.[5] Globally, it is estimated that the Earth's forests absorb as much as 30% of human-induced CO_2 emissions.[6]

Sustainably harvested wood products and materials also store atmospheric CO_2 long after they have been removed from a forest, with one cubic meter of wood capable of storing nearly a metric ton of CO_2[7]. In addition to carbon sequestration, forests also provide other benefits, such as clean water and wildlife habitat, recreational opportunities, and a source of living-wage jobs in rural communities. These attributes positively align with the UN's Sustainable Development Goals and contribute to advancing the UN's mission for a sustainable future for all. In light of these considerations, there is increasing awareness that well-managed forests are a critical component of any global climate change strategy.

CG uses scenario analyses to identify climate-related risks beginning at a broad country-level scale, narrowing down to a specific property, and then testing the impact of various risks to site suitability now and into the future. Factors analyzed to gauge climate risks include precipitation patterns, temperature fluctuations, the severity of weather events, the presence of pests or disease, and the annual average growth rates for commercial tree species. While many climate-related risks in forestry are mitigated through active management, during this iterative process CG analyzes both the potential positive and negative impacts associated with these risks to assess potential changes in net asset value. The following table illustrates climate risks evaluated, their impact on the forest, and our ability to mitigate the risks.

Climate Risk	Implication	CG Mitigants
Change in temperature	Increased fire danger	Property-specific fire plans; re-evaluate target regions/country for investment
Change in precipitation patterns	Changes in tree species range; increased drought and related fire risk	Vegetation suitability modeling and genetic tree improvement; re-evaluate target regions/country for investment
Frequency of extreme weather events	Loss of standing timber from wind events	Re-evaluate target regions; property-specific response plans; geographically diverse portfolio construction

(continued)

5 Estimated carbon storage of a 20-inch Douglas fir using the National Tree Benefit Calculator.
6 Y. Pan et al., "A Large and Persistent Carbon Sink in the World's Forests," *Science* (19 August 2011): 988–93.
7 Oregon Forest Resources Institute, "Forest to Frame" (2017).

Climate Risk	Implication	CG Mitigants
Presence of pests or disease	Early onset and increased frequency of individual tree mortality	Immediate treatment, which may include removal of effected trees to prevent further spread of pests or disease in the forest
Change in growth	Increased or decreased growth rates	Effects will vary by region, may influence planting stock decisions; re-evaluate forest growth model assumptions

Specific examples of how CG identifies opportunities and challenges in its investment process related to climate change include the following:

- Identifying afforestation opportunities that mitigate climate change by sequestering CO_2 emissions from the atmosphere in trees and soil, while offering many important co-benefits for communities, biodiversity, and soil and water quality.
- Protecting existing carbon stocks by minimizing impacts to carbon stored on the forest floor through tailored forest management practices.
- Enhancing forest carbon sequestration by replanting areas as soon as possible so the new forest will quickly begin removing CO_2 from the atmosphere.

The ability to quantify, evaluate, and report the year-over-year changes in the carbon footprint of a forest has the potential to influence the impacts an organization has on the environment, leading to increased transparency and more-informed business decisions. Incorporating climate change factors in its investment process not only mitigates climate-related risks; it also promotes and enhances the natural solutions forests provide. Understanding and measuring the comprehensive carbon stores of forests may lead to business decisions that improve carbon sequestration, a critical factor in addressing climate change.

Exhibit 25 shows the quarterly correlation between timber, farmland, and selected global equity and global bond indexes. In the 30-year period from 1990 through 2020, timber and farmland exhibited low correlation with traditional assets; the correlation of timber with global stocks and global bonds was 0.04 and 0.07, while the correlation of farmland with global stocks and global bonds was 0.15 and –0.04, respectively.

Exhibit 25 Historical Commodity Return Correlations, 1990–Q1 2020

NCREIF Data	Timber	Farmland
Global stocks	0.04	0.15
Global bonds	0.07	–0.04

Sources: Global stocks = MSCI ACWI; global bonds = Bloomberg Barclays Global Aggregate Index. Quarterly basis.

The United States has experienced three periods of elevated inflation as measured by the US Consumer Price Index since the early 1900s: 1915–1920, 1940–1951, and 1967–1981. In all three periods, farmland provided investors with meaningful inflation protection: +9.9% annualized farmland compound return versus +13.5% annualized inflation, +7.9% versus +5.9%, and +12.0% versus +7.6%, respectively.

Real estate return data pitfalls will be discussed in more detail in the next section, but note the inherent smoothing in quarterly versus monthly or daily data and the relative infrequency of transactions that make return projections problematic. After all, can one reliably or consistently compare a plot of prime farmland with a more marginal neighbor? How about the price of one piece of land that has generous water rights with another that is water restricted? As is the case with other alternative benchmarks (e.g., hedge fund indexes), productive land indexes may offer a good overall understanding but limited specific value.

KNOWLEDGE CHECK

Is the following statement true or false?

For the most appropriate instrument to invest in natural resources today, retail investors should focus on the stocks and bonds of companies producing in this sector.

A True, statement to support this option on why it's true.

B False, statement to support this option on why it's false.

Solution:

B is correct; the statement is false. Up to about 20 years ago, the only commonly available investment vehicles related to this asset class were indeed financial instruments (stocks and bonds). Rather than investing in the physical land and the products that come from it, investors concentrated on the companies that produced natural resources. Nowadays, however, the wide variety of direct investments via ETFs, limited partnerships, REITS, swaps, and futures opens the door for almost everyone to participate in these assets directly.

6.5 Instruments

To achieve the exposures they are after, investors require suitable instruments. Fortunately, there are many publicly traded and indexed choices, and there is still room for specialized value-added alternative mangers.

6.5.1 *Instruments: Commodities*

The majority of commodity investing is implemented through derivatives, and commodity futures are a popular choice. Physical commodities often generate unwelcome price opacity, tax obligations, and costs arising from storage, brokerage, and transportation, all of which may increase the attractiveness of standardized futures traded on transparent exchanges. Commodity derivatives include futures, forwards, options, and swaps. These contracts may be traded on exchanges or over the counter (OTC).

Futures contracts are obligations to buy or sell a specific amount of a given commodity at a fixed price, location, and date in the future. Futures contracts are exchange traded, are marked to market daily, and may or may not be settled with the delivery or receipt of the physical commodity at the end of the contract. This delivery obligation became dramatically important during the global financial crisis in 2008 and in the 2020 COVID-19 pandemic as demand collapsed, oil producers could not find buyers for their petroleum, and global storage filled suddenly. Even commodity-related ETFs were affected, forcing some to close and impose large losses on their investors. For futures contracts, counterparty risk is managed by the exchange and clearing broker. Commodity exposure can be achieved through means other than direct investment in commodities or commodity derivatives, including the following:

- Exchange-traded products (ETPs, either funds or notes) may be suitable for investors who are restricted to equity shares or are seeking the simplicity of trading through a standard brokerage account. ETPs may invest in commodities or commodity futures. For example, the SPDR Gold Shares exchange-traded fund attempts to track the price of physical gold by holding bullion in vaults. It owned just under $49 billion in gold bullion as of March 2020. ETPs may use leverage and may be long or short (also known as "inverse"). Similar to mutual funds or unit trusts, ETPs charge fees that are included in their expense ratios.

- Commodity exposure can also be obtained by investing in managed futures, also known as CTAs. However, in order to obtain pure commodity exposure, one would need to choose to invest in a managed futures fund focused solely on commodities because modern CTAs often invest in a variety of futures, including commodities, equities, fixed income, and foreign exchange. Such a fund might concentrate on a specific commodity, such as grains or livestock, or it may be broadly diversified to cover all commodities. Commodity-focused managed futures funds are unique (versus global macro) because there is a constant price tension between suppliers and consumers: High prices cripple demand (and thus lead to lower prices), and low prices shut in supply (and thus raise prices). This situation creates a unique balance that is not present in traditional asset classes, such as stocks and bonds.
- Funds that specialize in specific commodity sectors exist. For example, private energy partnerships, similar to private equity funds, are a popular way for institutions to gain exposure to the energy sector. Management fees can range from 1% to 3% of committed capital, with a lockup period of 10 years and extensions of 1- and 2-year periods. Publicly available energy mutual funds and unit trusts typically focus on the oil and gas sector and often act as fixed-income investments, paying out dividends from rents or capital gains. They may focus on upstream (drilling), midstream (refineries), or downstream (chemicals). Management fees for these funds are in line with those of other public equity managers and range from 0.4% to 1%.

6.5.2 *Instruments: Timberland and Farmland*

The primary investment vehicles for timber and farmland are investment funds, whether offered on the public markets, such as real estate investment trusts (REITs) in the United States, or administered privately through limited partnerships. Larger investors can consider direct investments if there are particular assets that have appeal. For example, Middle Eastern sovereign wealth funds have made investments in farmland in Africa and Southeast Asia. Owning physical farmland opens the door to a wider variety of foodstuffs: spices, nuts, fruits and vegetables—a much broader array than the corn, soy, and wheat offered by futures investment. However, there is limited price transparency or information to guide investment decisions without the assistance of sector specialists. The illiquidity of direct farm and timberland investments is also limiting.

PRACTICE QUESTIONS FOR LOS F

Question 1

A significant challenge to investing in timber is *most likely* its:

A high correlation with other asset classes.

B dependence on an international competitive context.

C return volatility compounded by financial market exposure.

Solution:

B is correct. A primary risk of timber is the international competitive landscape. Timber is a globally sold and consumed commodity subject to world trade interruptions. So the international context can be considered one of its major risk factors.

A is incorrect because timberland offers an income stream based on the sale of trees, wood, and other timber products that has not been highly correlated with other asset classes.

C is incorrect because investors are interested in timber because of its global nature (everyone requires shelter), the current income generated from the sale of the crop, inflation protection from holding the land, and its safe haven characteristics (it offers some insulation from financial market volatility).

Question 2

A characteristic of farmland strongly distinguishing it from timberland is its:

A commodity price-driven returns.

B inherent rigidity of production for output.

C value as an offset to other human activities.

Solution:

B is correct. Unlike timberland products, farm products must be harvested when ripe, so there is little flexibility in the production process. In contrast, timber (trees) can be grown and easily "stored" by simply not harvesting. This feature offers the flexibility of harvesting more trees when timber prices are up and delaying harvests when prices are down.

A is incorrect because just as a primary return driver for timberland is change in commodity price (of lumber from cut wood) in either the spot or futures price, farmland's returns are driven by agricultural commodity prices, with commodity futures contracts potentially combined with farmland holdings to generate an overall hedged return.

C is incorrect because for both farmland and timberland owned or leased for the benefit of the bounty each generates in the form of crops and more broadly timber, since these resources consume carbon as part of the plant life cycle, the considered value comes not just from the harvest but also from the offset to other human activities.

Question 3

Which of the following statements about commodity investing is invalid?

A Few commodity investors trade actual physical commodities.

B Commodity producers and consumers both hedge and speculate.

C Commodity indexes are based on the price of physical commodities.

Solution:

C is correct. Commodity indexes typically use the price of futures contracts on the commodities included in them rather than the prices of the physical commodities themselves in order to be transparent, investable, and replicable.

A is incorrect because trading in physical commodities is primarily limited to a smaller group of entities that are part of the physical supply chain. Thus, most commodity investors do not trade actual physical commodities but, rather, trade commodity derivatives.

B is incorrect because although supply chain participants use futures to hedge their forward purchases and sales of the physical commodities, those commodity producers and consumers nonetheless both hedge and speculate on commodity prices.

Question 4

An investor seeks a current income stream as a component of total return and desires an investment that historically has low correlation with other asset classes. The investment *most likely* to achieve the investor's goals is: (2020 Q10)

A timberland.

B collectibles.

C commodities.

Solution:

4. A is correct. Timberland offers an income stream based on the sale of timber products as a component of total return and has historically generated returns not highly correlated with other asset classes.

Question 5

If a commodity's forward curve is downward sloping and there is little or no convenience yield, the market is said to be in:

A backwardation.

B contango.

C equilibrium.

Solution:

5. B is correct. Contango is a condition in the futures markets in which the spot price is lower than the futures price, the forward curve is upward sloping, and there is little or no convenience yield. Backwardation is the opposite condition in the futures markets, where the spot price exceeds the futures price, the forward curve is downward sloping, and the convenience yield is high. Equilibrium is an economic term where supply is equal to demand.

7 REAL ESTATE

g explain investment characteristics of real estate

7.1 Overview of the Real Estate Market

Individuals and institutions buy real property for their own use, as an investment, or both. The residential, or housing, market is made up of individual single-family detached homes and multi-family attached units, which share at least one wall with another unit (e.g., condominiums, townhouses), owned by the resident. Commercial real estate includes primarily office buildings, retail shopping centers, and warehouses. In contrast to the owner-occupied market, rental properties are leased to tenants. A lease is a contract that conveys the use of the property from the owner (the landlord or *lessor*) to the tenant (*lessee*) for a predetermined period in exchange for compensation (rent). When residential real estate, be it single or multi-unit, is owned with the intention to let, lease, or rent the property in order to generate income, it is classified as commercial (i.e., income-producing) real estate.

Real estate investing is typically thought of as either direct or indirect ownership (equity investing) in real estate property, such as land and buildings. However, it also includes lending (debt investing) against real estate property. Loans secured by real estate are called mortgages. Investors can access real estate debt by lending, buying mortgages, or purchasing mortgage-backed securities with the property ultimately serving as collateral. The key reasons for investing in real estate include the following potential benefits:

- Competitive long-term total returns driven by both income generation and capital appreciation
- Multiple-year leases with fixed rents for some property types potentially providing stable income over many economic cycles
- Historically low correlations with other asset classes
- Inflation hedge if leases provide regular contractual rent step-ups or can be frequently marked to market

A title or deed represents real estate property ownership and covers building and land-use rights along with air, mineral, and surface rights. Titles can be purchased, leased, sold, mortgaged, or transferred together or separately, in whole or in part. Title searches are a crucial part of buyer and lender due diligence, ensuring the seller/borrower owns the property without any liens or other claims against the asset, such as from other owners, lenders, or investors or from the government for unpaid taxes.

Residential real estate is by far the largest sector of the real estate market by value and size. Savills World Research estimated in July 2018 that residential real estate accounted for more than 75% of global real estate values. Although the average value of a home is less than the average value of an office building, the aggregate space required to house people is much larger than the space required to accommodate office use and retail shopping. Homeownership rates are at least 80% in many of the largest countries by population, including China, India, Mexico, Japan, and Russia, and are high in other large countries, such as Brazil, at more than 74%, and the United States, with at least 64% homeownership. Germany is a notable exception, at approximately 52%, according to OECD.org.

The size of the professionally managed, institutional-quality global real estate market increased to US$9.6 trillion at the end of 2019, as shown in Exhibit 26, from US$6.8 trillion in 2013, representing a compound average annual growth rate of 5.8%. It should come as no surprise that the largest and most developed countries have the highest-valued real estate markets. Institutional-quality real estate generally consists of high-quality properties owned by institutional investors and high-net-worth individuals. It is typically of higher value than most individuals can afford. Smaller and out-of-date properties typically do not qualify as institutional quality.

Exhibit 26 Size of Professionally Managed Global Real Estate Market, 2018 (US$ billions)

Country or Region	Size	% of Total
United States	3,418.1	35.8%
Japan	881.4	9.2%
United Kingdom	745.5	7.8%

(continued)

Exhibit 26 (Continued)

Country or Region	Size	% of Total
China	592.2	6.2%
Germany	580.1	6.1%
France	441.2	4.6%
Hong Kong SAR	378.3	4.0%
Canada	361.0	3.8%
Australia	306.8	3.2%
Rest of world	1,848.4	19.3%
Total	9,553.0	100%

Source: MSCI Real Estate.

Real estate is different from other asset classes: the large capital investment required; the illiquidity; the fact that no two properties are identical in terms of location, tenant credit mix, lease term, age, and market demographics; and the necessarily fixed location of the asset. All these have important investment implications.

Furthermore, price discovery in the private market is opaque, historical prices may not reflect market conditions, transaction costs are high, and transaction activity may be limited in certain markets. It may be difficult for small investors to establish a diversified portfolio of wholly owned properties. Private market indexes are not investable, and property typically requires professional operational management.

Numerous other considerations will give investors pause before investing independently. Involvement in unfamiliar real estate markets calls for specialized knowledge of country, regional, and local market dynamics and the practices, regulations, and taxation that apply. Government regulations dictate what can be built or modified, with further complications coming from each locality's zoning and permit approval process, rules around ownership transfer, and possible rent controls. Local supply and demand dynamics may override wider market trends, complicating analysis. Many countries have limits on foreign investors with respect to property type and location, and they may require higher withholding taxes on foreign-owned property.

KNOWLEDGE CHECK

True or false: The largest sector of the real estate market by value and size is commercial real estate.

A True, statement to support this option on why it's true.

B False, statement to support this option on whey it's false.

Solution:

B is correct; the statement is false. Residential real estate is by far the largest sector of the real estate market by value and size. The residential debt market greatly exceeds commercial property debt because of the larger total value of residential properties combined with property owners' greater ability to use leverage—up to 80% of the property's value or more in some cases. In addition, home mortgages are subsidized in some markets, including government guarantees.

7.2 Characteristics: Forms of Real Estate Ownership

Real estate investing takes a variety of forms best illustrated with the four-quadrant model that arranges the quadrants by type of capital (debt or equity) and source of capital (private or public markets). Investors can choose to own all or part of a property's equity, either unlevered (100% equity) or with a levered approach that relies on debt as well as equity to finance the purchase. Alternatively, investors can gain exposure to real estate through debt ownership, either as a lender/originator or as a purchaser of debt instruments.

Banks and insurance companies are among the largest originators/owners of mortgage debt. The residential debt market greatly exceeds commercial property debt because the total value of residential properties is larger and property owners have a greater ability to use leverage—securing loans for up to 80% of the property's value or more in some cases. In addition, home mortgages are subsidized in some markets and enjoy government guarantees.

Exhibit 27 presents examples of the basic forms of real estate in each of the quadrants: public or private structures, debt or equity ownership.

Exhibit 27 Basic Forms of Real Estate Investments and Examples

	Debt	Equity
Private	■ Mortgages ■ Construction lending ■ Mezzanine debt	■ Direct ownership of real estate: ownership through sole ownership, joint ventures, separate accounts, or real estate limited partnerships ■ Indirect ownership via real estate funds ■ Private REITs
Public	■ MBS (residential and commercial) ■ Collateralized mortgage obligations ■ Mortgage REITs ■ ETFs that own securitized mortgage debt	■ Shares in real estate operating and development corporations ■ Listed REIT shares ■ Mutual funds ■ Index funds ■ ETFs

Within the basic forms, there can be many variations.

7.2.1 *Direct Real Estate Investing*

Direct private investing involves purchasing a property and originating debt for one's own account. Ownership can be free and clear, whereby the title to the property is transferred to the owner(s) unencumbered by any financing liens, such as from outstanding mortgages. Owners can also borrow from mortgage lenders to fund the equity acquisition. Debt investors can also use leverage. Initial purchase costs associated with direct ownership may include legal expenses, survey costs, engineering/environmental studies, and valuation (appraisal) fees. In addition, ongoing maintenance and refurbishment charges are also incurred. Additional debt closing costs are incurred when owners take out loans to fund their investments.

There are several benefits to owning real estate directly. One is control. Only the owner decides when to buy or sell, when and how much to spend on capital projects (subject to lease terms and regulatory requirements), whom to select as tenants based on credit quality preference and tenant mix, and what types of lease terms to offer. Another benefit relates to taxes. Property investors in the largest countries as ranked by the size of the institutional-quality real estate market can use non-cash property depreciation expenses and interest expense, with some limitations, to reduce taxable income and lower their income tax bills. In fact, real estate investors in a country that permits accelerated depreciation and interest expense deductions can reduce taxable income below zero in the early years of asset ownership, and losses can be carried forward to offset future income. Thus, a property investment can be cash-flow positive while generating accounting losses and deferring tax payments. If the tax losses do not reverse during the life of the asset, depreciation-recapture taxes can be triggered when the property is sold.

Major disadvantages to investing directly include the extensive time required to manage the property, the importance of local real estate market expertise to success, and the large capital requirements. Smaller investors bear additional risks from portfolio concentration.

The owner may choose to handle all aspects of investing in and operating the property, including property selection, asset management, property management, leasing, and administration. To address the complexity of owning and managing commercial real estate, investors hire advisers to perform any number of functions, including identifying investments, negotiating terms, performing due diligence, conducting operations, asset management, and eventual disposal. Asset management focuses on maximizing property returns by deciding when and how much to invest in the property and when to sell the property.

Many investors prefer to hire advisers or managers to manage the investors' direct real estate investment in what is called a *separate account*, which contains only the single investor's equity or perhaps a nominal stake from the adviser to help align interests. A separate account allows the investor to control the timing and value of acquisitions and dispositions and perhaps to make operating decisions as well. There is no commingling of capital with other investors.

Sometimes investors will form joint ventures with other investors to access real estate. Joint ventures are especially common when one party can contribute something of value that the other lacks, such as land, capital, development expertise, debt due diligence, or entrepreneurial talent. Joint ventures can be structured as general partnerships or, more commonly, as limited liability companies.

Numerous parties can pool their resources to acquire property, to develop or redevelop assets, or to lend capital. The equity stakes that investors contribute to pooled investment vehicles vary on the basis of their contribution of real assets, capital, entrepreneurial talent, and services. When structured as a limited partnership, there must by a general partner to manage the partnership and accept unlimited liability.

7.2.2 *Indirect Real Estate Investing*

Indirect investing provides access to the underlying real estate assets through a variety of pooled investment vehicles, which can be public or private, such as limited partnerships, mutual funds, corporate shares, REITs, and ETFs. Equity REITs own real estate equity, mortgage REITs own real estate mortgages and MBS, and hybrid REITs own both. Institutional investors and some high-net-worth investors can invest in MBS. Intermediaries facilitate the raising and pooling of capital and creating investable structures.

7.2.3 *Mortgages*

Mortgages represent passive investments in which the lender expects to receive a predefined stream of payments throughout the finite life of the mortgage. Mortgages may require full amortization or partial amortization with balloon payments due at maturity or may be interest only (IO). Borrowers can often choose among fixed-rate, floating-rate, and adjustable-rate options. Some of the floating- and variable-rate mortgages may have limitations, or caps, on how much the rate can change over a given period. For example, a borrower could apply for a mortgage that carries a fixed rate for five years, after which the rate resets to a predetermined spread over the lender's benchmark rate, with any rate change limited to 200 bps during the first year of the reset. Residential mortgages in developed markets usually have 15-, 25-, or 30-year maturities. Commercial property loans may be similarly long for long-term property owners. Other owners will take out 5-, 7-, or 10-year loans to correspond with the owners' anticipated holding period. Selling the property before the mortgage is due may result in prepayment penalties for the commercial borrower. If the borrower defaults on the loan, the lender may seek to take possession of the property. Investments may take place in the form of "whole" loans based on specific properties, typically by way of direct investment through private markets, or through participation in a pool of mortgage loans, typically by way of indirect investment in real estate through publicly traded securities, such as MBS.

REITs and partnerships carry fees and administrative overhead for managing the assets embedded in their valuations. Fee structures for private REITs can be similar to those for private equity funds, with investment management fees based on either committed capital or invested capital. These fees typically range from 1% to 2% of capital per annum. Funds also charge performance-based fees.

7.2.4 *Private Fund Investing Styles*

Capital committed to or invested by *real estate private equity funds* exceeded $900 billion in mid-2018. Most real estate private equity funds are structured as *infinite-life open-end funds*, which, like mutual funds, allow investors to contribute or redeem capital throughout the life of the fund. Investor subscriptions and redemptions are commonly accepted by the GP on a quarterly basis and are subject to the manager's ability to acquire or sell real estate or match investor redemption requests with pending investor subscriptions. Open-end funds generally offer exposure to *core* real estate, which is characterized by well-leased, high-quality institutional real estate in the best markets. Investors expect core real estate to deliver stable returns, primarily from income. Returns are typically driven by real estate beta, although managers can add value through better market selection, property management, and execution.

In addition to investing in core real estate, *core-plus* strategies will also accept slightly higher risks derived from *non-core* markets and sectors or properties with slightly more leasing risk. Non-core properties include large sectors with different risk profiles, such as hotels and nursing homes. Assets in secondary and tertiary markets and such niches as student housing, self-storage, and data centers are also considered non-core.

Investors seeking higher returns may also accept development, redevelopment, repositioning, and leasing risk. *Finite-life closed-end* funds are more commonly used for alpha- and beta-generating value-add and opportunistic investment styles. *Value-add* investments may require modest redevelopment or upgrades, the leasing of vacant space, or repositioning the underlying properties to earn a higher return than core properties. *Opportunistic* investing accepts the much higher risks of development, major redevelopment, repurposing of assets, taking on large vacancies, and speculating on significant improvement in market conditions. Closed-end fund managers will sell assets to realize the value created by management and lock in the investment's IRR. There are exceptions to the fund style/life/structure relationships—that is, core/infinite life/open end or opportunistic/finite life/closed end. On occasion, you will see core, closed-end, finite-life funds or open-end, infinite-life, value-add funds.

7.2.5 *REITs*

Real estate investment trusts are the preferred investment vehicles for owning income-producing real estate of both private and public investors. The main appeal of the REIT structure is the elimination of double corporate taxation: Corporations pay taxes on income, and then the dividend distributions of after-tax earnings are subsequently taxed at the shareholder's personal tax rate. REITs can avoid corporate income taxation by distributing dividends equal to 90%–100% of their taxable net rental income.

In the United States, taxable real estate corporations can elect REIT status by meeting a number of important criteria. Among the many rules, the company must invest at least 75% of its assets in real estate, cash, or US Treasuries and derive at least 75% of its income from rents on real property or real estate mortgages.

REIT and REIT-like structures in most countries are similar to those in the United States. Whereas REITs are a popular vehicle in the United States, other countries have structures separate from corporations for REITs that still permit dividend deductibility from the entities' income taxes. Primary differences relate to whether REITs can be internally managed, operated, and advised or are required to hire external managers and advisers, how much income the REIT must distribute (90%–100% is the standard), and whether the REIT is a corporation as in the United States or an independent structure. REITs that are self-managed and self-administered typically benefit from a better alignment of interests between management and shareholders than externally advised or managed companies.

Companies that have other sources of non-qualifying income or wish to retain earnings to reinvest in the business will choose an operating company structure. Note that the income taxes due for companies that can deduct depreciation and interest expense from operating income in calculating pretax income may not be high enough to accept the restrictions that come with REIT and REIT-like structures.

Publicly traded REITs address many of the disadvantages related to private real estate investing. Listed REITs provide investors with much greater liquidity, lower trading costs, and better transparency. Management is employed internally by the REIT in most countries rather than brought in as a separate organization on contract, which results in better alignment of interests with investors. Importantly, when an investor wants to modify its real estate holdings, the investor need only buy or sell REIT shares on listed markets instead of buying or selling real estate directly. The REIT is not forced to sell the company's underlying real estate the way open-end funds are when there are mass redemptions.

KNOWLEDGE CHECK

Fill in the blank: Compared with private real estate investing, publicly traded real estate investment trusts provide much greater ___________, lower ___________, and better ___________.

Choices of words: transparency, liquidity, trading costs, variety, happiness

Solution:

Compared with private real estate investing, publicly traded real estate investment trusts provide much greater <u>liquidity</u>, lower <u>trading costs</u>, and better <u>transparency</u>.

7.3 Characteristics: Real Estate Investment Categories

The majority of real estate property may be classified as either commercial or residential. In this reading, residential properties are defined narrowly to include only owner-occupied, single residences (often referred to as single-family residential property). Residential properties owned with the intention to let, lease, or rent them are classified as commercial. Commercial properties also include office, retail, industrial and warehouse, and hospitality (e.g., hotel and motel) properties. Commercial properties may also have mixed uses. Commercial properties generate returns from income (e.g., rent) and capital appreciation. Several factors will affect opportunities for capital appreciation, including development strategies, market conditions, and property-specific features.

7.3.1 *Residential Property*

For many individuals and families, real estate investment takes the form of direct equity investment (i.e., ownership) in a residence with the intent to occupy it. In other words, a home is purchased. Given the price of homes, most purchasers cannot pay 100% cash up front and must borrow funds to make the purchase. Most lenders require an equity contribution of at least 10%–20% of the property purchase price in countries with well-developed mortgage markets. Any appreciation (depreciation) in the value of the home increases (decreases) the owner's equity in the home and is magnified by mortgage leverage. In countries without well-developed mortgage markets, homebuyers must save for a much longer period and may have to ask family for assistance in order to raise enough money to buy a home.

Home loans may be held on the originator's balance sheet or securitized and offered to the financial markets. Securitization provides indirect debt investment opportunities in residential property to other investors via securitized debt products, such as residential mortgage-backed securities (RMBS).

7.3.2 *Commercial Real Estate*

Commercial property has traditionally been considered an appropriate direct investment, whether through equity or debt, for institutional funds and high-net-worth individuals with long time horizons and limited liquidity needs. This perception was primarily the result of the complexity, size, and relative illiquidity of the investments. Direct equity investing (i.e., ownership) is further complicated because commercial property requires active day-to-day management. The success of the equity investment is a function of a variety of factors, including how well the property is managed, general economic and specific real estate market conditions, and the extent and terms of any debt financing.

In order to provide direct debt financing, the lender (investor) will conduct financial analysis to determine the borrower's creditworthiness, to ensure the property can generate sufficient cash flow to service the debt, to estimate the property's value, and to evaluate economic conditions. The estimate of the property value is critical because the size of the loan relative to the property value, the loan-to-value ratio, determines the amount of risk held by the lender versus the borrower (equityholder). The borrower's equity in the property is an indicator of commitment to the success of the project, and it provides the lender with a cushion because the property is generally the sole collateral for the loan.

7.3.3 *REIT Investing*

As of 2019, REITs were listed on stock exchanges in 39 countries, and their combined market capitalization exceeded $1.6 trillion.[8] The risk and return characteristics of REITs depend on the type of investment they make. Mortgage REITs, which invest primarily in mortgages, are similar to fixed-income investments. Equity REITs, which invest primarily in commercial or residential properties and use leverage, are similar to direct equity investments in leveraged real estate.

---

8 www.reit.com/investing/global-real-estate-investment (accessed 12 August 2019).

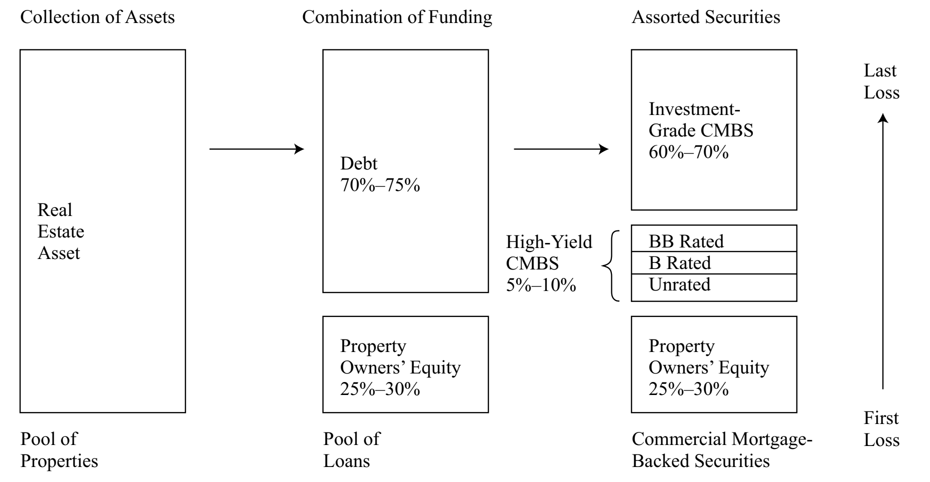
The business strategy for equity REITs is simple: Maximize property occupancy rates and rents in order to maximize income and dividends. Equity REITs, like other public companies, must report earnings per share based on net income as defined by generally accepted accounting principles (GAAP) or IFRS. Many report non-traditional measures, such as net asset value or variations of gross cash flow, to better estimate a company's dividend-paying ability, because non-cash depreciation expenses can be high for asset-intensive businesses.

7.3.4 *Mortgage-Backed Securities*

MBS structuring is based on the asset-backed securitization model of transforming illiquid assets (mortgages) into liquid securities and transferring risk from asset owners (banks, finance companies) to investors. An MBS issuer/originator forms a special purpose vehicle (SPV) to buy mortgages from lenders and other mortgage owners and use them to create a diversified mortgage pool. The MBS issuer assigns the incoming stream of mortgage interest income and principal payments to individual security tranches, which will be sold to investors. Each tranche is assigned a priority distribution ranking.

Exhibit 28 illustrates the basic process for creating commercial mortgage-backed securities (CMBS). On the right-hand side of the exhibit, the ranking of losses indicates the priority of claims against the real estate property. Risk-averse investors, primarily insurance companies, prefer the lowest-risk tranches, which are the first to receive interest and principal. Investors who choose the senior notes with the highest credit rating expect to earn a low return consistent with the senior tranche's low risk profile. Investors seeking the highest returns, which carry the highest risk, will invest in the lowest-rated, most junior securities, which are the last to receive interest and principal distributions. If mortgage defaults and losses are high, the lowest-ranked tranches bear the cost of the shortfall. The most junior tranche is referred to as the first-loss tranche.

Exhibit 28 CMBS Security Structure



MBS may be issued privately or publicly, and the securities are typically offered in $1,000 increments in the United States. MBS valuations are influenced by the underlying borrowers' behavior. When interest rates decline, fixed-income security prices usually rise, but borrowers are also likely to refinance their loans at a faster pace than before, resulting in the faster amortization of each MBS tranche, which leaves MBS investors to reinvest their principal at the lower rates. Conversely, when rates rise, not only do fixed-income instrument prices decline, but also property owner prepayments can also slow, lengthening the duration of most MBS tranches and contributing to further price weakness. These unusual traits are well understood by investors. MBS are often included in broad fixed-income indexes and in indexes that are used to indicate the performance of real estate investments.

7.4 Risk and Return Characteristics

7.4.1 *Real Estate Indexes*

There are a variety of indexes globally that are designed to measure total and component real estate returns for listed securities and non-listed investment vehicles. Listed REIT indexes are straightforward in that, like other listed equity indexes, security pricing is reported by the major exchanges. Total returns are calculated assuming dividends are reinvested in the index. Importantly, listed REIT indexes are investable when their constituent stocks are freely traded. The more frequently the shares trade, the more reliable the index. Indexes containing small-cap, closely held, and low-float real estate stocks may not be as representative of investor returns, and REIT indexes in general are not necessarily representative of the entire real estate universe because of different geographic concentrations, property types (residential, commercial, etc.), and the asset quality of the publicly traded REIT company portfolios. The National Association of Real Estate Investment Trusts (Nareit) in the United States, the European Public Real Estate Association (EPRA), and the Asian Public Real Estate Association (APREA) publish listed real estate company indexes for their respective regions and contribute to global indexes. Most of the leading for-profit index providers also offer a variety of REIT and real estate indexes.

As shown in Exhibit 29, REITs are the predominant structure among listed real estate companies in the United States. Outside the United States, there is an equal balance of REITs and non-REITs, as represented by the FTSE EPRA Nareit index series. Some other countries in which REITs make up most of the index include Japan, the United Kingdom, Australia, and Singapore. In contrast, non-REIT property companies make up most of the real estate indexes in Hong Kong SAR, Germany, and Sweden.

Exhibit 29 REITs Make Up 50% of the Property Index Outside the United States

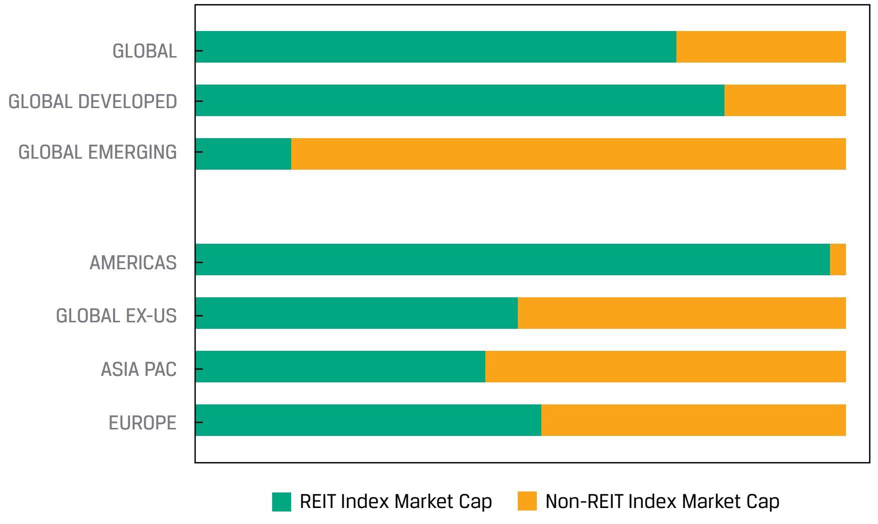


Exhibit 30 displays global and regional listed REIT returns. The table shows some disparity among regional returns and supports the importance of country-specific, regional, and local knowledge.

Exhibit 30 Historical Returns of Global REITs (Through 31 December 2019)

FTSE EPRA Nareit Developed Index Series, USD

Period	Global	North America	Asia Pacific	Europe
3 years	9.31%	7.76%	10.24%	13.05%
5 years	6.53%	6.63%	5.69%	7.40%
10 years	9.25%	11.39%	6.58%	8.66%
20 years	9.23%	11.06%	7.05%	9.38%
25 years	8.92%	11.13%	7.00%	8.88%
3-year volatility	8.24%	10.17%	9.05%	10.51%
10-year volatility	11.81%	13.05%	13.22%	15.16%
Dividend yield (Dec. 19)	3.83%	4.00%	3.60%	3.64%

Sources: FTSE, EPRA, NAREIT.

A variety of indexes measure private investment performance, and they may report on private fund performance, underlying property values, or property operating performance. Funds that track listed REIT indexes make public REIT performance investable, but the same instruments are not available for private indexes. Real estate values are reported by property owners, advisers, and fund managers, and many private investors do not report results, giving rise to selection bias. The private fund and property indexes vary by property selection, valuation methodology, and longevity. Although in many cases private index property values are based on third-party property appraisal and, therefore, offer a measure of independence from owner/manager estimates, appraisals are nonetheless backward looking, subject to the biases of the appraisers, and likely to offer a smoothed-out picture of actual market volatility.

Industry associations, practitioners, and academics have developed *repeat sales indexes* that are transaction based rather than appraisal based. When repeat sales take place, the changing property prices are measured and used to construct the index. These indexes also suffer from a sample selection bias because the properties that sell in each period vary and may not be representative of the larger market. Also, the properties that change hands are not a random sample and may be biased toward those that have increased or decreased in value, depending on economic conditions. The higher the number of sales, the more reliable and relevant the index.

In the United States, NCREIF, in conjunction with the Pension Real Estate Association (PREA), collects fund and property data, maintains numerous indexes, and disseminates the results. The NCREIF Property Index (NPI), which has data going back to 1977, measures property-level performance each quarter. NCREIF also produces open-end diversified core equity (OEDCE) fund, open-end (OE) fund, closed-end value-add (CEVA) fund, and transaction-based repeat sales indexes. The European Association for Investors in Non-Listed Real Estate (INREV) and the Asian Association for Investors in Non-Listed Real Estate (ANREV) produce similar indexes for their respective regions. These three organizations also contribute data from their respective non-listed real estate vehicle indexes, which are rolled up into the quarterly Global Real Estate Fund Index (GREFI).

A wide range of indexes for public stocks and private real estate funds and properties are published by such organizations as Cambridge Associates, Bloomberg, FTSE, MSCI, Real Capital Analytics, Savills, and S&P Dow Jones Indices.

Although investors can benefit from this variety and may well identify an index that is relevant to them, they should be aware of how these indexes are constructed and the inherent limitations that result. Investors should also be aware that real estate's apparent low volatility and low correlation with other asset classes may result from these limitations.

7.4.2 *Real Estate Investment Risks*

Real estate investments, like any investment, may fail to perform in accordance with expectations. Property values are variable in the face of national and global economic conditions, local conditions, and interest rate levels. There are risks around the ability of the fund management teams to select, finance, and manage real properties and to account for changes in government regulation. Management of the underlying properties themselves includes handling rentals and leases, controlling expenses, directing maintenance and improvements, and ultimately disposing of the property. Expenses may increase because of circumstances beyond management's control or be covered by insurance. Returns to both debt and equity investors in real estate depend to a large extent on the ability of the owners or their agents to successfully operate the underlying properties.

Investments in distressed properties and in property development are subject to greater risks than investments in properties that are in sound financial condition or enjoy stable operations. Property development is subject to special risks, including regulatory issues, construction delays, and cost overruns. Environmental regulation is one regulatory hurdle, as is the failure to receive zoning, occupancy, and other approvals and permits. Because the development and disposition period of such projects can be very lengthy, economic conditions may change. All these moving pieces may conspire to increase construction time or delay successful leases, which increases construction costs and reduces the level of rents relative to initial expectations. And there is also financing risk: Long-term financing with acceptable terms simply might not be available, forcing real estate acquisitions and developments to be financed with lines of credit or other forms of temporary financing. Financing problems with one property may delay or limit further development of the owner's other projects.

It is important to recognize that the vast majority of direct investors, private funds, and public companies pursue leverage to increase returns for their investors. Leverage magnifies the effects of both gains and losses for equity investors and increases the risks that the real estate owner or investor will be left with insufficient funds to make interest payments or repay the debt at maturity. In that case, the lender or debt investor will receive less than the entire interest and principal balances. As the loan-to-value ratio increases and interest coverage weakens, the probability of default increases.

KNOWLEDGE CHECK

Fill in the blanks: Adding real estate to a portfolio has been demonstrated to increase portfolio ___________ and reduce portfolio ___________.

Solution:

Adding real estate to a portfolio has been demonstrated to increase portfolio diversification and reduce portfolio risk.

7.5 Diversification Benefits

Many investors find the real estate sector attractive for providing high, steady current income. Although broad stock market gains are derived mostly from long-term capital appreciation, more than half of the returns from private core and listed real estate are derived from income. Throughout a market cycle, income is a more consistent source of return than capital appreciation, and hence, it is lower risk. The source of that consistency stems from the underlying medium- to long-term property leases: The longer the lease the more predictable the income. And the better the credit quality of a property's tenants, the more reliable the rents.

Are REITs equity, debt, or something else? In line with our categorization of all the assets in this reading, many investors view REITs as alternative investments in the broader real estate category, and others see REITs as a fixed-income/equity hybrid, adding exposure simply for the incremental yield pickup over both high-grade bonds and equities. Others still have a different focus: Some institutional investors use their public equity department to analyze their REIT investments rather than engaging their real estate or alternative asset teams.

Real estate's moderate correlation with other asset classes adds to its attractiveness. Exhibit 31 lists the correlation of REITs with equity and debt securities during the 20-year period ended March 2020. The market capitalization of most REITs is on the smaller side, with the largest among them being notable exceptions.

Exhibit 31 REIT Allocation Can Improve Portfolio Diversification

Index	Correlation with the FTSE Nareit All Equity REITs Index
S&P 500	0.57
Russell 2000 Index	0.64
ICE BofAML Corp./Gov't.	0.19
ICE BofAML MBS	0.09
Bloomberg Barclays US Agg./Credit/Corp. High Yield	0.61

Sources: Nareit, using data from FactSet and Intercontinental Exchange (ICE).

There are periods when equity REIT correlations with other securities are elevated, and they are at their highest during steep market downturns, such as the 2007–08 financial crisis. These correlations remained high during the post-crisis recovery, which lifted the value of most asset classes.

Nevertheless, the diversification benefit persists, and including real estate has been demonstrated to reduce portfolio risk. Oxford Economics studied the expected performance of listed European real estate as an asset class, comparing it with equities, fixed income, and commodities.[9] "A substantial allocation to listed real estate," the authors concluded, "does enhance the risk-return characteristics of a multi-asset portfolio." They went on to recommend larger allocations to real estate for European investors.

9 Listed Real Estate in A Multi-Asset Portfolio: A European Perspective, pg,. 2, Oxford Economics, EPRA, September 2019.

Exhibit 32 contains a summary of the performance and correlations of major asset classes and shows that European investors would benefit from real estate allocations.

Exhibit 32 Total Returns by Asset Class, Jan. 1999–May 2019

	Total Return (CAGR)	Standard Deviation	Sharpe Ratio	Correlation with Listed Real Estate (Annual Returns)
Small-cap equities	8.2%	22.9%	0.47	0.60
Large-cap equities	2.2	21.3	0.21	0.55
Government bonds	3.8	3.9	1.02	0.15
Investment-grade corporate bonds	4.3	5.5	0.84	0.41
High-yield corporate bonds	7.1	18.0	0.45	0.44
Diversified commodities	1.8	18.9	0.19	0.14
Listed real estate	8.4	22.7	0.49	1.00

Source: Data sources are as follows: small-cap equities, S&P Europe SmallCap index; large-cap equities, STOXX Europe 50 Index; government bonds, Bloomberg Barclays Pan-European Aggregate Government A Index; investment-grade corporate bonds, Bloomberg Barclays Pan-European Aggregate Corporate Index; high-yield corporate bonds, Bloomberg Barclays Pan-European High Yield Total Return Index; diversified commodities, Bloomberg Commodity Index; listed real estate, FTSE EPRA Nareit Developed Europe Index.

Whether listed real estate behaves like stocks or private real estate is a matter of ongoing debate. We know that listed REITs are priced continuously whereas private real estate is appraised perhaps once a year, and there are other indications that their correlation is weak. Listed equity investors will discount future cash flows, and appraisers place heavy emphasis on current market conditions and recent trends.

CEM Benchmarking, based in Toronto, has performed several studies looking at the role of listed and unlisted real estate in the portfolio of defined contribution pension funds. CEM Benchmarking also concludes that real estate improves the risk–return profile of funds diversified across equities, fixed income, and alternatives. CEM analyzed the performance of each asset class over 12 years, from 2005 to 2016, across €2 trillion of AUM at the end of 2016. The data covered 36% of European pension assets and provided a comprehensive view about how real estate performs relative to the pension funds' other assets. Exhibit 33 presents the asset correlations for Dutch and UK pension funds, which are among the deepest pension fund markets with real estate concentrations.

Exhibit 33 Correlations between Aggregate Asset Class Net Returns for Dutch and UK Pension Funds

Correlation Matrix	Public Equity	Private Equity	Fixed Income	Hedge Funds	Listed Real Estate	Unlisted Real Estate	Unlisted Infrastructure
Public equity	1.00						
Private equity	0.93	1.00					
Fixed income	0.12	0.06	1.00				
Hedge funds	0.91	0.77	0.16	1.00			

(continued)

Exhibit 33 (Continued)

Correlation Matrix	Public Equity	Private Equity	Fixed Income	Hedge Funds	Listed Real Estate	Unlisted Real Estate	Unlisted Infrastructure
Listed real estate	0.86	0.83	0.28	0.81	1.00		
Unlisted real estate	0.76	0.86	0.37	0.64	0.88	1.00	
Unlisted infrastructure	0.73	0.67	0.56	0.62	0.66	0.67	1.00

Notes: The period spans 2005–2016. Unlisted real estate and private equity are adjusted for lagged and smoothed reporting. Actual, as-reported correlations would appear higher and standard deviations, lower.

CEM Benchmarking conducted another study addressing some of the differences between listed and unlisted real estate performance by comparing the returns, volatility, correlations, and Sharpe ratios for various private real estate investing styles. The study covered more than 200 US public and private defined benefit pension funds over 20 years through 2017. Unlisted real estate's correlation with listed real estate was 0.91, which supports the view that listed REITs behave more like stocks in the short term and more like real estate in the long term. The outperformance of listed real estate was primarily due to fees and carry charged by external managers in the unlisted markets.

PRACTICE QUESTIONS FOR LOS G

Question 1

The majority of real estate property may be classified as either:

A debt or equity.

B commercial or residential.

C direct ownership or indirect ownership.

Solution:

B is correct. The majority of real estate property may be classified as either commercial or residential.

Question 2

Which of the following relates to a benefit when owning real estate directly?

A Taxes

B Capital requirements

C Portfolio concentration

Solution:

A is correct. When owning real estate directly, there is a benefit related to taxes. The owner can use property non-cash depreciation expenses to reduce taxable income and lower the current income tax bill. In fact, accelerated depreciation and interest expense can reduce taxable income below zero in the early years of asset ownership, and losses can be carried forward to offset future income. Thus, a property investment can be cash-flow positive while generating accounting losses and deferring tax payments. If the tax losses do not reverse during the life of the asset, depreciation-recapture taxes can be triggered when the property is sold.

B is incorrect because the large capital requirement is a major disadvantage of investing directly in real estate.

C is incorrect because a disadvantage for smaller investors who own real estate directly is that they bear the risk of portfolio concentration.

Question 3

Which of the following statements is true regarding mortgage-backed securities?

A Insurance companies prefer the first-loss tranche.

B When interest rates rise, prepayments will likely accelerate.

C When interest rates fall, the low-risk senior tranche will amortize more quickly.

Solution:

C is correct. When interest rates decline, borrowers are likely to refinance their loans at a faster pace than before, resulting in faster amortization of each MBS tranche, including the senior tranche, which is the lowest-risk tranche.

A is incorrect because risk-averse investors, primarily insurance companies, prefer the lowest-risk tranches, which are the first to receive interest and principal. The junior-most tranche is referred to as the first-loss tranche. It is the highest-risk tranche and is the last to receive interest and principal distributions.

B is incorrect because when interest rates rise, prepayments will likely slow down, lengthening the duration of most MBS tranches. Prepayments will likely increase when interest rates decline, because borrowers are likely to refinance their loans at a faster pace.

Question 4

Which of the following statements is true for REITs?

A According to GAAP, equity REITs are exempt from reporting earnings per share.

B Though equity REIT correlations with other asset classes are typically moderate, they are highest during steep market downturns.

C The REIT corporation pays taxes on income, and the REIT shareholder pays taxes on the REIT's dividend distribution of after-tax earnings.

Solution:

B is correct. Real estate investments, including REITs, provide important portfolio benefits due to moderate correlation with other asset classes. However, there are periods when equity REIT correlations with other securities are high, and their correlations are highest during steep market downturns.

A is incorrect because equity REITs, like other public companies, must report earnings per share based on net income as defined by GAAP or IFRS.

C is incorrect because REITs can avoid this double taxation. A REIT can avoid corporate income taxation by distributing dividends equal to 90%–100% of its taxable net rental income. This ability to avoid double taxation is the main appeal of the REIT structure.

Question 5

What is the most significant drawback of a repeat sales index to measure returns to real estate? (2020 Q14)

A Sample selection bias

B Understatement of volatility

C Reliance on subjective appraisals

Solution:

A is correct. A repeat sales index uses the changes in price of repeat sales properties to construct the index. Sample selection bias is a significant drawback because the properties that sell in each period vary and may not be representative of the overall market the index is meant to cover. The properties that transact are not a random sample and may be biased toward properties that changed in value. Understated volatility and reliance on subjective appraisals by experts are drawbacks of an appraisal index.

Question 6

As the loan-to-value ratio increases for a real estate investment, risk *most likely* increases for: (2020 Q19)

A debt investors only.

B equity investors only.

C both debt and equity investors.

Solution:

C is correct. The higher the loan-to-value ratio, the higher leverage is for a real estate investment, which increases the risk to both debt and equity investors.

8 INFRASTRUCTURE

h explain investment characteristics of infrastructure

8.1 Introduction and Overview

The assets underlying infrastructure investments are real, capital intensive, and long lived. These assets are intended for public use, and they provide essential services—for instance, airports, health care facilities, and power plants. Most infrastructure assets are financed, owned, and operated by governments. A 2017 World Bank/PPIAF study of infrastructure spending in developing countries reported that 83% came from public investment,[10] but more and more infrastructure is being financed privately, with the increasing use by local, regional, and national governments of public–private partnerships. PPPs are typically defined as a long-term contractual relationship between the public sector and the private sector for the purpose of having the private sector deliver a project or service traditionally provided by the public sector.

10 World Bank Group and Public-Private Infrastructure Advisory Facility, "Who Sponsors Infrastructure Projects? Disentangling Public and Private Contributions" (2019).

Allocations to infrastructure investments have increased not only because of increasing interest by investors (demand-side growth) but also because of governments providing more investment opportunities (supply-side growth) as they expand the financing of infrastructure assets and the privatization of services. By the end of 2019, according to Preqin, global infrastructure funds had raised more than US$98 billion for infrastructure projects, a year-over-year increase of approximately 15% and up from US$30 billion in 2012, with assets under management of US$582 billion (as of the end of June 2019),[11] reflecting what is commonly considered a growth market. Indeed, Global Infrastructure Hub (a G–20 initiative) and Oxford Economics forecast an infrastructure investment gap of US$15 trillion in 2040, based on the US$79 trillion presently forecast to be spent. (A further $3.5 trillion is required by 2030 to meet the United Nations' Sustainable Development Goals for electricity and water).[12]

Infrastructure investors may intend to lease the assets back to the government, to sell newly constructed assets to the government, or to hold and operate the assets until they reach operational maturity or perhaps for even longer. Exhibit 34 shows the characteristics of a typical infrastructure investment.

Exhibit 34 Typical Infrastructure Investment Characteristics

From an investment perspective, if the assets are being held and operated, the relatively inelastic demand for the assets and services is advantageous; regulation and the high costs of the assets create high barriers to entry, which give the provider of the services a strong competitive position. (Regulation, conversely, deters adverse monopolistic behavior from the public perspective.) Maintenance, asset replacement, and operating costs are factored into the pricing of the services.

Investors expect these assets to generate stable long-term cash flows that adjust for economic growth and inflation. Well-defined contractual structures that allocate the risk and responsibilities of asset delivery, service provision, and legal and financial obligations contribute to such stability.

11 Preqin, "2020 Preqin Global Infrastructure Report" (4 February 2020).
12 Global Infrastructure Hub and Oxford Economics, "Global Infrastructure Outlook" (2017).

These structural aspects allow for relatively high levels of leverage, which is particularly important for funding given the typically high up-front capital investment. Leverage enhances investor returns, and with debt commonly issued on a non-recourse basis, the exposure of investors to their investment is capped.

An increasingly important theme of infrastructure investing is the incorporation of ESG criteria, whether through the application of technology (e.g., renewable energy sources), consideration of the project's impact on the surrounding environment, or how the criteria inform the business relationships and governance approach. All key infrastructure stakeholders are increasingly supporting this theme: the government/public procuring authority, by incorporating environmental and social impact assessments in the planning and permit processes; financiers, through, for instance, the application of the Equator Principles as a condition for advancing loan facilities; and investors, who consciously use ESG criteria to choose investments. The Equator Principles represent a risk management and due diligence framework that has been adopted by financial institutions globally to take account of environmental and social risk in projects.

8.2 Description

8.2.1 *Categories of Infrastructure Investments*

Infrastructure investments are frequently categorized on the basis of the underlying assets. The broadest categorization organizes investments into economic and social infrastructure assets.

Economic infrastructure assets support economic activity and include three broad types of assets: transportation assets, information and communication technology (ICT) assets, and utility and energy assets:

- *Transportation assets* include roads, bridges, tunnels, airports, seaports, and heavy and light/urban railway systems. Income will usually be linked to demand based on traffic, airport, and seaport charges, tolls, and rail fares and hence is deemed to carry market risk.
- *ICT assets* include infrastructure that stores, broadcasts, and transmits information or data, such as telecommunication towers and data centers.
- *Utility and energy assets* generate power and produce potable water; they transmit, store, and distribute gas, water, and electricity; and they treat solid waste.

Utility assets increasingly encompass environmentally sustainable development, with a greater focus on renewable technologies, including solar, wind, and waste-to-energy power generation. (The increasing rejection of less environmentally viable fuel sources, such as coal, is part of the story.) Other energy assets may encompass downstream oil and gas infrastructure, such as pipelines and LNG terminals, and also natural resource assets, such as mining assets. The income earned from utility assets may also carry demand risk as buyers' energy and natural resources needs fluctuate. Alternatively, utilities can institute "take-or-pay" arrangements that lock buyers into minimum purchases whether they need the supply or not. Buyers usually have recourse if the utility falls short on performance, delivering supplies that are late or of inferior quality.

Social infrastructure assets are directed toward human activities and include such assets as educational, health care, social housing, and correctional facilities, with the focus on providing, operating, and maintaining the asset infrastructure. The relevant services that are administered through those facilities, meanwhile, be they medical or schooling, are usually provided separately by the public authority or by a private service provider contracted by the public authority. In some countries, this model has been extended to other public infrastructure, such as courthouses and government and municipal buildings. Income from social infrastructure is typically derived from a type of lease payment that depends on assuring availability (often referred to as availability payments) and on managing and maintaining the asset according to predefined standards. For instance, an availability payment may be reduced or voided if a hospital operating room is not available because of an electromechanical fault.

Infrastructure investments may also be categorized by the underlying asset's stage of development. Investing in infrastructure assets *that are to be constructed* is generally referred to as **greenfield investment**. The intent may be to lease or sell the assets to the government after construction or to either hold and operate the assets over the long term or hold and operate them for a shorter period until operational maturity and then sell them to new investors, thus ensuring capital appreciation that reflects the construction and commissioning risk. Greenfield investors typically invest alongside strategic investors or developers who specialize in developing the underlying assets. Investing in *existing* infrastructure assets may be referred to as **brownfield investment**. Perhaps the assets are owned by a government that wants to privatize them, lease them out, or sell them and lease them back; or they may be owned by greenfield investors seeking to realize the value of their investment through a sale. Typically, some of the assets' financial and operating history is available. Brownfield investors may include strategic investors who specialize in the operation of the underlying assets, but particularly in the case of privatizations, there will be financial investors involved who focus on the long-term stable returns.

Finally, infrastructure may be categorized by location. Infrastructure investments are available globally, and the geographic location of the underlying assets will inform the political and macroeconomic risks, particularly in light of the government's relationship to the assets.

Risks and expected returns may differ on the basis of the underlying asset's category, stage of development, and geographic location. How the investment is actualized—its form—also affects the risks and expected returns.

KNOWLEDGE CHECK

Describe one of the three ways by which an infrastructure investment is typically categorized.

Solution:

Infrastructure assets are frequently characterized on the basis of (1) underlying assets, (2) the underlying asset's state of development, and (3) the geographic location on the underlying assets.

The first category comprises economic assets supporting economic activity, such as transportation and energy/utility assets. It also includes social infrastructure assets, which enable public services directed toward human activities, such as education, health care, and housing.

The second category refers to the relevant stages of the infrastructure asset's life cycle. These include a greenfield investment for a newly created asset and a brownfield investment for an established asset.

The third category highlights the specific place (local, regional, national) associated with the government entity directly involved with the assets.

8.2.2 *Forms of Infrastructure Investments*

As is the case with real estate investments, infrastructure investments take a variety of forms. The choice affects liquidity, cash flow, and income streams. Investing directly in the underlying assets provides control and the opportunity to capture full value. It entails a large investment, however, and results in concentration and liquidity risks while the assets have to be managed and operated. Hence, direct investment often happens in consortia, with strategic partners who are better placed to manage certain risks (e.g., operational) or with other financial/institutional investors to limit individual concentration risk. However, most investors invest indirectly.

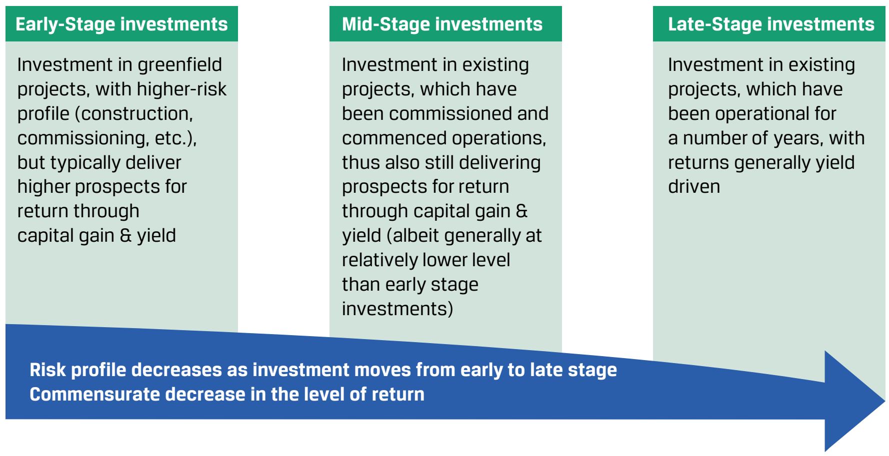
Indirect investment vehicles include infrastructure funds (which are similar in structure to private equity funds and may be closed end or open end), infrastructure ETFs, and company shares. Different asset class strategies, including debt and mezzanine investment, are offered, but the vast majority of investors are focused on equity instruments. According to IJInvestor, equity is the preferred instrument for 77% of fund investors, with 10% looking for mixed strategies and 9% looking for debt and mezzanine strategies.[13]

Investors concerned about liquidity and diversification may choose publicly traded infrastructure securities or master limited partnerships. Publicly traded infrastructure securities provide the benefits not only of liquidity but also of reasonable fees, transparent governance, market prices, and the ability to diversify among underlying assets. An investor should be aware, however, that publicly traded infrastructure securities represent a small segment of the infrastructure investment universe and tend to be clustered in certain asset categories. Master limited partnerships (MLPs) trade on exchanges and are pass-through entities, similar to REITs. They also share with REITs applicable regulations that vary among countries, and income is passed through to the investors for taxation; MLPs generally distribute most free cash flow to their investors. Typically, the GP manages the partnership, receives a fee, and holds a small partnership interest, with LPs owning the remaining majority partnership interest.

8.3 Risk and Return Characteristics

The lowest-risk infrastructure investments have more stable cash flows and higher dividend payout ratios. They typically enjoy fewer growth opportunities, however, and lower expected returns. For example, an MLP with an investment in a brownfield asset that is being leased back to a government represents a low-risk infrastructure investment, as does an investment in assets with a history of steady cash flows, such as certain established toll roads. An investment in a fund that is building a new power plant without any operating history (a greenfield investment) is riskier.

Exhibit 35 Typical Infrastructure Investment Risk–Return Profile



13 IJInvestor, "IJInvestor Funds & Investors Report - Q1 2020" (17 April).

There is a risk that infrastructure revenues diverge meaningfully from expectations, leverage creates financing risk, and operational risk and construction risk are ever present. As for regulatory risk, because the projects involve essential services, governments usually regulate the investments with strictures on the sale of the underlying assets, on operations and service quality, and on prices and profit margins. Global infrastructure investing introduces additional uncertainty, such as currency risk, political risk, and profit-repatriation risk. Hence, the fund manager must ensure these risks are appropriately managed and mitigated through insurance, financial instruments, and whatever other mechanisms are available.

Exhibit 36 Typical Infrastructure Risk Management

PERFORMANCE RISKS	**Demand/ volume risk**	Enter into take-or-pay arrangements, where payments are based upon the availability rather than the use of an asset
	Operational risk	Enter into operation and maintenance contracts with reputable and experienced operators who are incentivized to meet or exceed the minimum contractual performance standards through an appropriate regime
	Construction risk	Enter into fixed-price date-certain contracts with reputable and experienced contractors to minimize construction cost overruns and delays
STRUCTURAL RISKS	**Financing/ interest rate risk**	Enter into interest rate swaps
	Regulatory risk	If regulation is by contract (e.g., PPP agreement), ensure clear, unambiguous provisions that govern such risks; if regulation is defined and enforced by a regulatory body, then ensure due diligence is undertaken on that body
	Political risk	Purchase political risk insurance for protection, as appropriate, against events such as expropriation, political violence, sovereign debt default, etc.
	Currency risk	Ensure PPP agreement incorporates appropriate mechanisms to permit an adjustment to tariffs when there are material foreign exchange fluctuations
	Tax/profit repatriation risks	Ensure PPP agreement incorporates appropriate mechanisms to permit an adjustment to tariffs and/or other compensation when there are significant changes in law/regulations affecting tax and profit repatriation

Returns, of course, depend on such factors as investment type. Exhibit 37 provides an illustrative range of returns for a private infrastructure fund according to different risk profiles, noting that returns may vary at an individual investment/asset level depending on its specific characteristics.

Exhibit 37 Private Infrastructure Fund Illustrative Target Returns

Higher-Risk Profile	Medium-Risk Profile	Lower-Risk Profile
Greenfield projects without guarantees of demand upon completion—e.g., variable electricity prices, uncertain traffic on roads and through ports	Mostly brownfield assets (with some capital expenditure requirements) and some greenfield assets (with limited construction and demand risk)	Brownfield assets with mitigated risks—e.g., fully constructed with contracted/regulated revenues
Located in OECD countries and emerging markets	Located primarily in OECD countries	Located in the most stable OECD countries
High weighting to capital appreciation	Mix of yield and capital appreciation	High weighting to current yield
Target equity returns of 14%+	**Target equity returns of 10%–12%**	**Target equity returns of 6%–8%**

Note: Target equity returns are net of fees.
Source: Cambridge Associates, "Digging In: Assessing the Private Infrastructure Opportunity Today," Research Note (April 2017).

Most infrastructure funds gravitate toward the medium- and lower-risk profiles. Rolling one-year-horizon returns have been around 10% annually since 2016.[14]

In summary, infrastructure investments generally provide investors with stable, long-term capital growth and cash distributions, according to the risks assumed, and these risks tend to be relatively well defined. Preqin maintains a return series of private funds investing in infrastructure deals. Standard & Poor's, FTSE Group, and other firms also publish indexes of publicly traded infrastructure companies.

KNOWLEDGE CHECK

Fill in the blank by choosing from among the following options:

___________ is an infrastructure investment characteristic *most likely* valuable to investors aiming to sell newly constructed assets to the government.

A "Strategically important"

B "Monopolistic and regulated"

C "Significant capital investment"

Solution:

A is the correct answer because the related priority will probably increase the demand of the public buyer to effectively provide essential services to its citizens. B would be more advantageous for investors holding and operating an asset and charging fees to the buyer, with the inelastic demand supporting pricing and regulations increasing the barriers to entry that improve the competitive position. C is more of a challenge than a benefit to the investor, requiring greater funding capability and potentially higher financial risks.

8.4 Diversification Benefits

We have established that investors expect infrastructure assets to generate stable long-term cash flows that adjust for economic growth and inflation. Investors may also expect capital appreciation, depending on the type and timing of their investment.

14 "2020 Preqin Global Infrastructure Report."

Investing in infrastructure may allow for the addition of an income stream, for further portfolio diversification by adding an asset class with typically low correlation with existing investments, and for some protection against inflation. Low exposure to short-term GDP growth issues may also be a factor.

Infrastructure has proven to be relatively resilient to recent swings in the equity markets. According to Preqin, from December 2017 to December 2018 as equity returns fell from 22.8% to –9.6% and hedge fund returns fell from 12.2% to –3.05%, infrastructure returns dropped only from 11.4% to 9.6%.[15] Such resilience is a key factor for many institutional investors with an eye on portfolio construction.

Exhibit 38 provides an illustrative example of the low correlation of infrastructure—and in particular, unlisted Australian infrastructure—with other asset classes over the 10-year period through June 2019.

Exhibit 38 Quarterly Return Correlations for Selected Asset Classes, 2009–2019

	Global Treasury Bonds	Global Equities	Australian Equities	Global REITs	Listed Infrastructure	Unlisted Australian Infrastructure
Global Treasury bonds	1.00					
Global equities	–0.49	1.00				
Australian equities	–0.36	0.89	1.00			
Global REITs	0.18	0.64	0.70	1.00		
Listed infrastructure	–0.12	0.81	0.75	0.68	1.00	
Unlisted Australian infrastructure	–0.08	0.06	–0.05	–0.21	0.13	1.00

Notes: All data are for the period 31 March 2009–30 June 2019, except for the FTSE EPRA NAREIT Developed Rental Net index (Global REITs), where data series began in May 2009.

- Global Treasury bonds: Bloomberg Barclays Global Treasury GDP Index
- Global equities: MSCI World Net Accumulation Index
- Australian equities: S&P/ASX 200
- Global REITs: FTSE EPRA NAREIT Developed Rental Net index
- Listed infrastructure: Dow Jones Brookfield Global Infrastructure Net Accumulation Index
- Unlisted infrastructure: MSCI Australian Unlisted Infrastructure Index

Source: John Julian, "The Pros of Infrastructure Investment in a Lower-for-Longer Environment," AMP Capital (October 2019). www.ampcapital.com/au/en/insights-hub/articles/2019/october/the-pros-of-infrastructure-investment-in-a-lower-for-longer-environment.

Because these investments are placed in long-lived assets, infrastructure may better match the longer-term liability structure of certain investors, such as pension funds, superannuation schemes, and life insurance companies. It also suits the longer-term horizon of sovereign wealth funds, which tend to make the largest allocations to this asset class—around 5%–6% of total AUM, according to Preqin.[16]

15 "2020 Preqin Global Infrastructure Report."
16 "2020 Preqin Global Infrastructure Report."

KNOWLEDGE CHECK

To highlight typical risk management approaches on infrastructure projects, draw connecting lines to match a particular risk type with its appropriate mitigating mechanism:

Infrastructure Risk Type	Offsetting Mechanism
Construction	Incentivized maintenance contract
Operational	Insurance against debt default
Demand/volume	PPP agreement
Currency	Take-or-pay arrangement
Political	Fixed-price date-certain contract

Solution:

The correct matches are as follows:

Infrastructure Risk Type	Offsetting Mechanism
Construction	Incentivized maintenance contract
Operational	Insurance against debt default
Demand/volume	PPP agreement
Currency	Take-or-pay arrangement
Political	Fixed-price date-certain contract

PRACTICE QUESTIONS FOR LOS H

Question 1

Compared with direct investment in infrastructure, publicly traded infrastructure securities are characterized by:

A higher concentration risk.

B more transparent governance.

C greater control over the infrastructure assets.

Solution

B is correct. Publicly traded infrastructure securities, which include shares of companies, exchange-traded funds, and listed funds that invest in infrastructure, provide the benefits of transparent governance, liquidity, reasonable fees, market prices, and the ability to diversify among underlying assets. Direct investment in infrastructure involves a large capital investment in any single project, resulting in high concentration risks. Direct investment in infrastructure provides control over the assets and the opportunity to capture the assets' full value.

Question 2

Which of the following forms of infrastructure investment is the most liquid?

A An unlisted infrastructure mutual fund

B A direct investment in a greenfield project

C An exchange-traded MLP

Solution:

C is correct. A publicly traded infrastructure security, such as an exchange-traded MLP, provides the benefit of liquidity.

Question 3

An investor chooses to invest in a brownfield, rather than a greenfield, infrastructure project. The investor is *most likely* motivated by:

A growth opportunities.

B predictable cash flows.

C higher expected returns.

Solution:

B is correct. A brownfield investment is an investment in an existing infrastructure asset, which is more likely to have a history of steady cash flows compared with that of a greenfield investment. Growth opportunities and returns are expected to be lower for brownfield investments, which are less risky than greenfield investments.

Question 4

The privatization of an existing hospital is best described as:

A a greenfield investment.

B a brownfield investment.

C an economic infrastructure investment.

Solution:

B is correct. Investing in an existing infrastructure asset with the intent to privatize, lease, or sell and lease back the asset is referred to as a brownfield investment. An economic infrastructure asset supports economic activity and includes such assets as transportation and utility assets. Hospitals are social infrastructure assets, which are focused on human activities.

Question 5

Risks in infrastructure investing are *most likely* greatest when the project involves:

A construction of infrastructure assets.

B investment in existing infrastructure assets.

C investing in assets that will be leased back to a government.

Solution:

A is correct. Infrastructure projects involving construction have more risk than investments in existing assets with a demonstrated cash flow or investments in assets that are expected to generate regular cash flows.

9 ISSUES IN PERFORMANCE APPRAISAL

i. describe issues in performance appraisal of alternative investments

9.1 Overview of Performance Appraisal for Alternative Investments

Portfolio managers invest in one of two basic ways to achieve returns: passively or actively. Passive investments focus on index or asset class coverage. Portfolios comprising real estate, commodity, and infrastructure instruments may be passively managed to provide exposure to these alternative asset classes. However, alternative investments are generally actively managed, aiming to achieve added portfolio benefits, but typically at a higher net cost for such active management.

Investors frequently look to alternative investments for diversification and a chance to earn relatively high returns on a risk-adjusted basis. They also value low correlation and a more risk-neutral source of alpha.

Evaluating an alternative investment can be a subtle qualitative exercise, one that depends on the initial objectives of the investor, as opposed to a purely quantitative, one-size-fits-all exercise. Much of the nuance revolves around not just the total net return created by an alternative investment but also the path and volatility (drawdown risk) required to create the total return and how an alternative investment fits into and benefits a larger portfolio of assets—in other words, its portfolio-level correlation benefit.

However, there is often a naive attraction to alternative investments based on their expected returns, which neglects to consider the atypical risks they present—risks we can examine on both a standalone and portfolio basis:

- Limited transparency
- Often low portfolio-level liquidity
- The active use of leverage (and, at times, derivatives)
- General strategy and product complexity
- Mark-to-market issues for a portfolio with niche specialized products
- Limited redemption availability and portfolio pressures from eventual redemption activities
- The general challenge of manager differentiation and diversification
- The fee drag associated with alternative investments, which can be non-trivial

In this section, we examine some of these issues in the context of applying traditional performance measurement tools to alternative investments, which is sometimes done incorrectly. In Section 10, we will delve into the impact of fee structures on alternative asset evaluation.

9.2 Common Approaches to Performance Appraisal and Application Challenges

9.2.1 *Sharpe Ratio*

The Sharpe ratio is probably the first basic intuitive metric that some people use to evaluate an alternative investment. It is often prominently displayed in marketing materials. The single biggest flaw, however, in a dependence on the Sharpe ratio is probably the underlying assumption of normally distributed returns.

In light of our list of atypical risks—leverage, illiquidity, mark-to-market smoothing, poor portfolio transparency—return profiles of alternative investments tend to be asymmetric and skewed, making the Sharpe ratio a less-than-ideal performance measure. For non-normal return distributions with significant skewness (fat tails in one direction or the other) and kurtosis (a measure of whether the data are heavy tailed or light tailed relative to a normal distribution), volatility is not a perfect measure of dispersion. Although still widely cited, the Sharpe ratio may not, therefore, be a good risk-adjusted performance measurement to rely on.

Given that caveat and taken by itself as a simple starting point, an attractive Sharpe ratio for an alternative strategy might be deemed to be anywhere between 1.0 and 2.0. Getting twice the return per unit of risk is generally quite attractive and hard to achieve; getting only a single unit of return for a unit of risk (a Sharpe ratio of 1.0) might be deemed acceptable but not overwhelmingly compelling, unless the strategy in question also has negative correlation attributes compared with other strategy areas in a portfolio, as with macro or CTA managers. In comparison, as of the end of 2019, the long-only S&P 500 equity index delivered an average annual return of around 10% over the past 32 years but with a relatively high annual standard deviation around the mean, 15%, yielding a Sharpe ratio close to 0.66. Alternative asset managers generally endeavor to offer a far superior risk–reward profile.

Ironically, if a Sharpe ratio for an alternative manager is too high—say, 3.0–4.0—extra caution in the evaluation of that manager may be warranted. These types of Sharpe ratios are typically only available if some degree of *return smoothing* is taking place. Illiquid securities that are hard to value and coupon return streams that might one day suddenly come to an end (sometimes along with a sudden loss of principal, as with a structured bond product that extinguishes itself or a short option strategy) can produce very high Sharpe ratios for a period of time that may deceive and then ultimately disappoint naive investors.

The ultimate test for a portfolio is really upon liquidation, but investment value at that point is frequently at odds with how accounting standards require alternative asset positions to be carried on the books. Furthermore, neither expected liquidation value nor reported value may account for the tail-risk aspects in an unfortunate macro environment.

The nature of the positions that a portfolio holds must be considered: Is the Sharpe ratio even the relevant metric?

9.2.2 *Sortino Ratio*

A second, arguably more valuable, quantitative measure of performance is return relative to downside volatility—the Sortino ratio. However, the Sharpe and Sortino risk measures alone still do not take into account the correlation of alternative assets with the traditional assets that their inclusion in the portfolio may be intended to hedge.

We have established that alternative investments' diversifying potential is part of the motivation for investing in them: Investors perceive an opportunity to improve the risk–return relationship in the portfolio context. Given the historical return, volatility, and correlation profiles of alternative investments, combining a portfolio of alternative investments with a portfolio of traditional investments should improve the overall portfolio's risk–return profile. Doing so can increase the risk-adjusted return of the overall portfolio because of potentially higher returns to the portfolio and a less-than-perfect correlation with traditional investments.

It is key for investors to consider how an alternative manager's investment choices will correlate with traditional assets and how the assets in question affect drawdown options and liquidity.

9.2.3 *Treynor Ratio and Correlation Behavior*

Alternative assets are generally deemed more valuable when they have a lower beta to traditional assets, and this is where the Treynor ratio proves useful: It is a measure of the excess average return of an investment relative to its beta to a relevant benchmark, such as a broad equity index.

The lower the beta of the alternative asset, the higher the Treynor ratio will be, and all else being equal, when comparing alternative investment possibilities, an asset with a higher Treynor ratio will be deemed more attractive than an asset with a low Treynor ratio.

The main limitation of the Treynor ratio is that it is based on historical beta data that may change in the future. The ratio also becomes less meaningful if the beta of the alternative asset to its systematic benchmark is negative, which is certainly possible for some alternative assets.

The Treynor ratio is the first ratio we have covered that encompasses the return of an alternative asset relative to its expected systematic risk. Examining an alternative asset's correlation coefficient with the overall market and with the other assets in the portfolio will also be instructive. All else being equal, less correlated alternative assets are generally deemed more attractive for portfolio diversification purposes than more correlated assets. It is nevertheless important to note that the expected diversification benefits from alternative investments are not always realized, even when they are most needed. Correlations between risky investments increase during periods of market stress and can approach 1.0 during financial crises.

For now, one easy proxy for measuring return relative to risk for alternative investments—and a far simpler calculation to observe and calculate—is to analyze the performance record for average return relative to the worst drawdown loss. A maximum drawdown (MDD) is the maximum observed loss from a peak to a trough of a portfolio, before a new peak is attained. The **Calmar ratio** is a comparison of the average annual compounded return to this maximum drawdown risk. The higher (lower) the Calmar ratio, the better (worse) an alternative asset performed on a risk-adjusted basis over a specified period of time. The Calmar ratio is typically calculated using the prior three years of performance, and it thus adjusts over time. Variations of the ratio exist: The **MAR ratio** uses a full investment history and the average drawdown. Both ratios help address the left-tailed return profile that sometimes characterizes alternative assets.

Other ratios are applied to specific alternative strategies and require more granular disclosure from investment managers. Some quantitative managers refer to the percentage of profitable trades (or "batting average") and their "slugging percentage," the magnitude of the gains from winning trades divided by the losses from losing trades. A manager who makes outsized gains on winning trades while incurring small losses on losing trades is worth considering.

KNOWLEDGE CHECK

True or false: The Sharpe ratio measures the amount of risk per unit of return.

A True, statement to support this option on why it's true.

B False, statement to support this option on why it's false.

Solution:

B is correct; the statement is false. The Sharpe, Sortino, and Treynor ratios are risk-adjusted performance measurements. They are measures of return per unit of risk.

KNOWLEDGE CHECK

Is the following statement true or false? The Sharpe and Sortino ratios share the same denominator.

A True, statement to support this option on why it's true.

B False, statement to support this option on why it's false.

Solution:

B is correct; the statement is false. The Sharpe and Sortino ratios share the same numerator—average annualized return net of the risk free-rate. Their denominators are different. The denominator for the Sharpe ratio is standard deviation of returns, and the denominator for the Sortino ratio is downside deviation of returns—a semi-deviation measure of volatility only during periods of loss for an alternative investment.

KNOWLEDGE CHECK

Which of the following performance measures uses beta as the risk measure?

A Sharpe ratio

B Treynor ratio

C Sortino ratio

Solution:

The correct answer is B. Both Sharpe Ratio and Sortino Ratio use standard deviation as a risk measure. Treynor ratio uses beta as a risk measure.

9.3 Private Equity and Real Estate Performance Evaluation

Unlike liquid hedge fund strategies, alternative investments that involve private equity and real estate tend to have variable cash flows.

Private equity investments generally involve an initial capital commitment, but actual capital flows often lag that commitment because capital "calls" are staggered over substantive periods of time. Private equity returns are frequently described in terms of the J-curve effect: a substantive initial capital commitment promise, followed by high initial fee drag (calculated on total committed capital, not the capital actually called), followed by the identification of longer-term growth or turnaround opportunities, and an eventual positive return (the investor hopes) when the staggered returns of the fund are realized. The component investments mature and are sold at various times, and the partnership finally closes—but usually only after six to eight years (often with two-year extensions allowed at the discretion of the general partner or after a vote of approval by a majority of limited partners). The line representing the return changes from downward sloping to positive and then to energetically upward sloping later in the investment's life (if all goes well).

The real estate pathway starts with initial property purchases, followed at times by substantive cash outlays for improvements, followed by instances of accounting depreciation that can influence after-tax performance, followed typically by the receipt of rents and then an eventual property sale (often at a long-term tax-advantaged tax rate).

The measurement of success in both instances depends far more on the timing and magnitude of cash flows in and out of the investments, and these are often hard to standardize, let alone always properly anticipate. Given the long time horizon, the application of different tax treatments can have a non-trivial impact on after-tax investment returns.

As a general rule, the best way to start evaluating such investments is with the IRR of the respective cash flows into an investment and the timing thereof, versus the magnitude and the timing of the cash flows returned by the investment (inclusive of tax benefits).

In an independent, fixed-life private equity fund, the decisions to raise money, take money in the form of capital calls, and distribute proceeds are all at the discretion of the private equity manager. Timing of cash flows is an important part of the investment decision process. The private equity manager should thus be rewarded or penalized for the results of those timing decisions, and the calculation of an IRR is a key metric for doing so.

Although the determination of an IRR involves certain assumptions about a financing rate to use for outgoing cash flows (typically a weighted average cost of capital) and a reinvestment rate to use for incoming cash flows (which must be assumed and may or may not actually be earned), the IRR is the key metric used to assess longer-term alternative investments in the private equity and real estate worlds.

Because of this complexity, a shortcut methodology often used by both private equity and real estate managers involves simply citing a **multiple of invested capital (MOIC)**, or money multiple, on total paid-in capital. Here, one simply measures the total value of all distributions and residual asset values (assets that may still be awaiting their ultimate sale) relative to an initial total investment. Although this shortcut valuation completely ignores the timing of cash flows, it is easier to calculate, and it is intuitively easier to understand when someone says they received two or three times their initial investment. But how long it takes to realize this value does matter. A 2× return on one's initial investment would be phenomenal if the return were collected over two years but far less compelling if it took 15 years to realize.

In general, because private equity and real estate investing involve longer holding periods, there is less emphasis on evaluating them in terms of shorter-term portfolio correlation benefits. After a private equity fund has fully drawn in its monetary commitments, interim accounting values for a private equity partnership become less critical for a period of time because no incoming or outgoing cash flows may immediately hinge on such valuations. During this "middle period" in the life of a private equity fund, accounting values may not always be particularly reflective of the future potential realizations (and hence the expected returns) of the fund. It is not that the value of the investments is not actually rising and falling in the face of economic influences; rather, accounting conventions simply leave longer-lived investments marked at their initial cost for some time or make only modest adjustments to carrying value until clearer impairments or realization events take place.

Ironically, many investors gravitate to private equity *because* they are not easily marked to market and can deliver somewhat smoothed returns over time—with less short-term mark-to-market angst. Although most private equity managers are conservative in their marks (awaiting actual realization events), a lagged mark-to-market process can at times offer a false sense of success, diminishing short-term investment worry but subsequently delivering the occasional back-end-loaded investment disappointment. Private equity interim valuations and real estate appraisals can certainly be inaccurate or skewed at times despite the best efforts of auditors to present a fair valuation.

The lagging impact of shorter-term economic events on the interim accounting valuations of these strategies makes them appear more resilient and less correlated than they really are. A more realistic picture may emerge when premature portfolio liquidations are forced on managers. The lack of transparency around such investments and the slowness to mark them to market can be incorrectly construed by investors as an overall lack of volatility.

Along this imperfect path and in an effort to benchmark how longer-term investments may be faring, private equity and real estate managers are generally judged by where they fall in terms of a *quartile ranking*, which depicts their performance against a cohort of peer investment vehicles constructed with similar investment attributes and funded around the same time, or what is often referred to as the same vintage year.

Real estate managers are also often judged by the **cap rate** being earned on their properties, which is simply the annual rent actually being earned (net of any vacancies) divided by the price originally paid for the property. This approach ignores the true values of properties should they need to be sold.

The large latitude around carried valuations for both private equity and real estate strategies makes any application of shorter-term risk metrics highly inappropriate.

Instead, the probability of permanent impairment on an investment is likely the best risk measure. Along these lines, an admittedly backward-looking metric of some interest and potential value that many private equity managers refer to is their historic **capital loss ratio**, defined as the percentage of capital in deals that have been realized below cost, net of any recovered proceeds, divided by total invested capital. Because markets have been generally buoyant for the past 30 years, investors can overestimate the true resilience of private equity and real estate. Having money tied up in longer-lived investments, for 7 to 10 years, should conceptually earn an illiquidity premium but may easily result in return disappointment given the wrong economic environment.

9.4 Hedge Funds: Leverage, Illiquidity, and Redemption Terms

9.4.1 *Leverage*

Hedge funds may use leverage to obtain higher returns on their investments. Leverage has the effect of magnifying gains and losses because it allows for taking a larger position relative to the capital committed. Hedge funds leverage their portfolios by using derivatives or borrowing capital from **prime brokers**, negotiating with them to establish margin requirements, interest, and fees in advance of trading. The hedge fund deposits cash or other collateral into a margin account with the prime broker, and the prime broker essentially lends the hedge fund the shares, bonds, or derivatives to make additional investments. The margin account represents the hedge fund's net equity in its positions. The minimum margin required depends on the riskiness of the investment portfolio and the creditworthiness of the hedge fund.

Leverage is a large part of the reason that some hedge funds either earn larger-than-normal returns or suffer significant losses. If the margin account or the hedge fund's equity in a position declines below a certain level, the lender initiates a margin call and requests that the hedge fund put up more collateral. An inability to meet margin calls can have the effect of magnifying or locking in losses because the hedge fund may have to liquidate (close) the losing position. This liquidation can lead to further losses if the order size is sufficiently large to move the security's market price before the fund can sufficiently eliminate the position.

Under normal conditions, the application of leverage may be necessary for yielding meaningful returns from given quantitative, arbitrage, or relative value strategies. But with added leverage comes increased risk. For example, in August 2007, a sudden cascade of problems arose from the application of leverage: so-called unattractive stocks rallied sharply over a 10-day period of short covering while stocks deemed fundamentally "attractive" declined as quantitative managers were forced to sell them. Many managers positioning themselves as market neutral and "safe" quickly lost 20%–25% of their portfolio value. The tail-risk aspect of such strategies was revealed to be the leverage and the "crowding" phenomenon of so many market participants holding similar positions that needed to be downsized when the market moved. Leverage can cause left-tailed events when mechanical downsizing and crowd psychology impacts come into play. Market watchers said that the source of the 2007 upheaval was no more than the butterfly effect of one levered quantitative manager after another deciding to de-lever at the same time.

The application of leverage to a strategy may not be revealed by studying the Sharpe ratio, the Sortino ratio, or another financial performance metric. Instead, it may underpin the strategy undetected. Understanding the impact of leverage on a portfolio is less about what the track record has been and more about *how* that track record was created. Analysts evaluating the alternative asset space should, therefore, scrutinize the returns of any alternative asset that relies on high leverage.

9.4.2 *Illiquidity and Potential Redemption Pressures*

A second qualitative issue for many hedge funds is the manner in which portfolios are marked to market. This issue is less important for traditional long–short managers trading only publicly traded equities, but it still can cause problems.

Consider the long–short equity manager involved in thinly traded small-capitalization stocks that see only a few thousand shares traded daily. Perhaps the manager was able at one point to source a block of stock from a retiring firm founder in order to establish the long exposure. An outside fund administrator uses a daily quoted price to value these shares, but the manager knows there is little chance of actually liquidating all the shares at that price.

This problem will be worse for a convertible bond manager, a credit-oriented manager, or a structured products manager who faces wider bid–offer spreads and deals in some securities that are particularly illiquid or that trade only "by appointment" (in other words, securities that trade so infrequently an appointment is almost necessary).

Proper valuations are important for calculating performance and meeting potential redemptions without incurring undue transaction costs for liquidating exposures. The frequency with which alternative assets are valued and how they are valued vary among funds. Hedge funds are generally valued by the manager internally, and these valuations are confirmed by an outside administrator on a daily or perhaps weekly basis; performance is reported to investors by the administrator on a monthly or quarterly basis. The valuation may use market values or estimated values of the underlying positions. When market prices or quotes are used for valuation, funds may differ in which price or quote they use (bid price, ask price, average quote, or median quote). A common practice is to use an average of the bid and the ask. A more conservative and theoretically more accurate approach is to use bid prices for long positions and ask prices for short positions because these are more realistic prices at which the positions could be closed.

In some instances, the underlying positions may be in highly illiquid or even non-traded investments, and since such securities may have no reliable market values, it becomes necessary to estimate values. Starting in 2006, with later amendments and modifications, GAAP accounting rules in the United States created a methodology that involves the categorization of investments into three buckets: Level 1, 2, and 3 asset pricing. Level 1 assets involve situations where an exchange-traded, publicly traded price is available and is mandated to be used for valuation purposes. When such pricing is not available, outside broker quotes, or Level 2 values, may be relied on. As a final recourse, when such broker quotes are deemed either unavailable or unreliable, assets may be computed using only internal models—a process referred to as Level 3 asset pricing. No matter the model used by a manager in such circumstances, it should be independently tested, benchmarked, and calibrated to industry-accepted standards to ensure a consistency of approach. Because of the potential for conflicts of interest when applying estimates of value, hedge funds must develop procedures for in-house valuation, communicate these procedures to clients, and adhere to them.

Notwithstanding best practice, the very nature of assets that can be valued only on a "mark-to-model" basis can and should be a concern for the alternative asset investor. A model may reflect an imperfect theoretical valuation and not a true liquidation value. The illiquid nature of these assets means that estimates, rather than observable transaction prices, may well have factored into any valuation. As a result, returns may be smoothed or overstated and the volatility of returns, understated. Any investor relying on a risk metric such as a Sharpe ratio for such illiquid strategy situations is basically involved in self-deception as to the true drawdown risk of the strategy. As a generalized statement, any investment vehicle that is heavily involved with Level 3–priced assets deserves increased scrutiny and due diligence.

EXAMPLE 1

Hedge Fund Valuation

A hedge fund with a market-neutral strategy restricts its investment universe to domestic publicly traded equity securities that are actively traded on an exchange or between over-the-counter brokers. In calculating net asset value, the fund is most likely to use which of the following to value underlying positions?

A Exchange last-trade pricing and/or averaged quotes of any available over-the-counter bid–offer spreads

B Average quotes adjusted for liquidity

C Bid price for shorts and ask price for longs

Solution:

A is correct. The fund is most likely to use exchange-traded last-trade pricing (Level 1 pricing) or averaged quotes of publicly available bid–offer spreads (Level 2 pricing). The securities are actively traded, so no liquidity adjustment is required.

If the fund uses bid–ask prices, it will use ask prices for shorts and bid prices for longs; these are the prices at which the positions are closed.

Understanding and evaluating "tail events"—low-probability, high-severity instances of stress—is an important yet extraordinarily difficult aspect of the risk management process, particularly when dealing with illiquid securities. Stress testing and scenario analysis are often used as a complement to other risk measures to develop a better understanding of a portfolio's potential loss under both normal and stressed market conditions. Stress testing involves estimating losses under extremely unfavorable conditions.

Another factor that can lock in or magnify losses for hedge funds is investor redemptions. Redemptions frequently occur when a hedge fund is performing poorly. Redemptions may require the hedge fund manager to liquidate some positions and potentially receive particularly disadvantageous prices when forced to do so by redemption pressures, while also incurring transaction costs. Funds sometimes charge redemption fees (typically payable to the remaining investors) to discourage redemption and to offset the transaction costs for remaining investors. Notice periods provide an opportunity for the hedge fund manager to liquidate a position in an orderly fashion without magnifying the losses. Lockup periods—time periods when investors cannot withdraw their capital—provide the hedge fund manager the required time to implement and potentially realize a strategy's expected results. If the hedge fund receives a drawdown request shortly after a new investment, the lockup period forces the investors who made the request to stay in the fund for a period of time rather than be allowed to immediately withdraw. A hedge fund's ability to demand a long lockup period while raising a significant amount of investment capital depends a great deal on the reputation of either the firm or the hedge fund manager. Funds of hedge funds may offer more redemption flexibility than is afforded to direct investors in hedge funds because of special redemption arrangements with the underlying hedge fund managers, the maintenance of added cash reserves, access to temporary bridge-loan financing, or the simple avoidance of less liquid hedge fund strategies.

Ideally, redemption terms should be designed to match the expected liquidity of the assets being invested in, but even with careful planning, an initial drawdown can turn into something far more serious when it involves illiquid and obscure assets. These left-tailed loss events are not easily modeled for hedge funds.

EXAMPLE 2

Effect of Redemption

A European credit hedge fund has a very short redemption notice period—one week—because the fund's managers believe it invests in highly liquid asset classes and is market neutral. The fund has a small number of holdings that represent a significant portion of the outstanding issue of each holding. The fund's lockup period has expired. Unfortunately, in one particular month, because of the downgrades of two large holdings, the hedge fund has a drawdown (decline in NAV) of more than 10%. The declines in value of the two holdings result in margin calls from their prime broker, and the drawdown results in requests to redeem 50% of total partnership interests. The combined requests are *most likely* to:

A force the hedge fund to liquidate or unwind 50% of its positions in an orderly fashion throughout the week.

B have little effect on the prices received when liquidating the positions because the hedge fund has a week before the partnership interests are redeemed.

C result in a forced liquidation, which is likely to further drive down prices and result in ongoing pressures on the hedge fund as it tries to convince nervous investors to remain in the fund.

Solution:

C is correct. One week may not be enough time to unwind such a large portion of the fund's positions in an orderly fashion that also does not further drive down prices. A downgrade is not likely to have a temporary effect, so even if other non-losing positions are liquidated to meet the redemption requests, it is unlikely that the two large holdings will return to previous or higher values in short order. Also, the hedge fund may have a week to satisfy the requests for redemptions, but the margin call must be met immediately. Overall, sudden redemptions at the fund level can have a cascading negative impact on a fund.

The previous discussion applies mostly to liquid alternative asset strategies, principally hedge funds. In the world of private equity and real estate alternative assets, other methodologies are used to measure relative performance. Here, we find more issues with lagged and smoothed pricing.

It is important to note that although the ratios we have considered are among the best performance and risk tools available, they can still lead to inappropriate conclusions, as shown in the following example.

EXAMPLE 3

Steamboat Structured Products LLC is a manager that specializes in the purchase and sale of residential mortgage-backed securities, sometimes hedged with other put option and short equity index exposures. For the most part, its strategy is geared to earn an attractive monthly mortgage payment on as low of a loan-to-value ratio of a secured property as possible. Steamboat managers are generally very good at sourcing such investments.

However, Steamboat's managers discover that in the wake of the 2008 financial crisis, many small pieces of "odd-lot" RMBS paper are being sold by financial institutions across the United States, sometimes with face values of just \$200,000–\$400,000. This paper is often offered at a 10% or 15% discount to the price a more sizable and significant "round-lot" block of \$1 million–\$2 million of the same type of paper might trade.

Steamboat managers know that their outside administrator will place only a single round-lot valuation on each security identifier (often referred to as a CUSIP) in their portfolio, and thus they can't resist starting to buy as many of these odd-lot offerings as possible. When they do, their administrator does indeed immediately mark them higher, which, as Steamboat continues this practice month after month, creates a lovely track record of constant "trading profits" on top of their natural coupon income. To the greatest extent possible, Steamboat hopes to create odd-lot "matchers" to eventually aggregate its exposures into larger tradable blocks. But what happens instead is that the firm ends up with 2,000 small-position line items in its book and a glorious track record.

This is a story of an illiquid asset class that is naturally hard to trade and mark. Steamboat's Sharpe ratio, Sortino ratio, and Calmar ratio all look stronger than they really should be, mostly because of accounting conventions. If Steamboat were ever forced to sell its portfolio, the odd-lot discount earned as trading revenue would largely disappear and be replaced by the true liquidation values that Steamboat would find in the market from others for its accumulated odd-lot portfolio.

Performance ratio analysis must, therefore, be discounted when dealing with illiquid securities as described here. Although not necessarily fraudulent, standard accounting practices can be purposefully gamed.

PRACTICE QUESTIONS FOR LOS I

Question 1

The Sharpe ratio is a less-than-ideal performance measure for alternative investments because:

A it uses a semi-deviation measure of volatility.

B returns of alternative assets are not normally distributed.

C alternative assets exhibit low correlation with traditional asset classes.

Solution:

B is correct. The Sharpe ratio assumes normally distributed returns. However, alternative assets tend to have non-normal return distributions with significant skewness (fat tails in one direction or the other) and kurtosis (sharper peak than a normal distribution has, with fatter tails). Therefore, the Sharpe ratio may not be a good risk-adjusted performance measure to rely on for alternative investments.

A is incorrect because the Sharpe ratio does not use a semi-deviation measure of volatility; it uses standard deviation. The Sortino ratio uses a semi-deviation measure of volatility. Further, the use of semi-deviation instead of standard deviation actually makes the Sortino ratio a more attractive measure of alternative asset performance than the Sharpe ratio.

C is incorrect because correlation does not enter into the calculation of the Sharpe ratio. However, it is true that alternative assets can have low correlations with other asset classes. In contrast to the Sharpe ratio, the Treynor ratio incorporates the beta of the alternative asset relative to a benchmark, which is conceptually similar to correlation.

Question 2

Which of the following is true regarding private equity performance calculations?

A The money multiple calculation relies on the amount and timing of cash flows.

B The IRR calculation involves the assumption of two rates.

C Because private equity funds have low volatility, accounting conventions allow them to use a lagged mark-to-market process.

Solution:

B is correct. The determination of an IRR involves certain assumptions about a financing rate to use for outgoing cash flows (typically a weighted average cost of capital) and a reinvestment rate assumption to make on incoming cash flows (which must be assumed and may or may not actually be earned).

A is incorrect because the money multiple calculation completely ignores the timing of cash flows.

C is incorrect because it is somewhat of a reversal of cause and effect: Private equity (PE) funds can appear to have low volatility because of the lag in their mark-to-market process. It's not that PE investments don't actually rise and fall behind the scenes with economic influences, but accounting conventions may simply leave longer-lived investments marked at their initial cost for some time or with only modest adjustments to such carrying value until known impairments or realization events begin to transpire. Also, because PE funds are not easily marked to market, their returns can appear somewhat smoothed, making them appear more resilient and less correlated with other assets than they really are. The slowness to re-mark them can unfortunately be confused by investors as an overall lack of volatility.

Question 3

Which is *not* true of mark-to-model valuations?

A Return volatility may be understated.

B Returns may be smooth and overstated.

C A calibrated model will produce a reliable liquidation value.

Solution:

C is correct. It is not true that a calibrated model will produce a reliable liquidation value in a mark-to-model valuation. The need to use a model for valuation arises when an asset is so illiquid that there are not reliable market values available. A model may reflect only an imperfect theoretical valuation, not a true liquidation value, should liquidation become necessary. The illiquid nature of alternative assets means that estimates, rather than observable transaction prices, may have been used for valuation purposes.

A and B are not correct because they are both true statements.

Question 4

An analyst wanting to assess the downside risk of an alternative investment is *least likely* to use the investment's: (2020 Q34)

A Sortino ratio.

B value at risk (VaR).

C standard deviation of returns.

Solution:

C is correct. Downside risk measures focus on the left side of the return distribution curve, where losses occur. The standard deviation of returns assumes that returns are normally distributed. Many alternative investments do not exhibit close-to-normal distributions of returns, which is a crucial assumption for the validity of a standard deviation as a comprehensive risk measure. Assuming normal probability distributions when calculating these measures will lead to an underestimation of downside risk for a negatively skewed distribution. Both the Sortino ratio and the VaR measure are measures of downside risk.

10 CALCULATING FEES AND RETURNS

j calculate and interpret returns of alternative investments on both before-fee and after-fee bases

Now that we have considered alternative assets, their risks and rewards, their investment characteristics, and headline strategies, performing a few calculations will take us close to the real-world evaluations made by investment managers and the investors who hire them.

10.1 Alternative Asset Fee Structures and Terms

EXAMPLE 4

Incentive Fees Relative to Waterfall Types

A PE fund invests \$15 million in a nascent luxury yacht manufacturer and \$17 million in a new casino venture. The yacht manufacturer generates a \$9 million profit when the company is acquired by a larger competitor, but the casino venture turns out to be a flop when its state licensing is eventually denied and it generates a \$10 million loss. If the manager's carried interest incentive fee is

20% of the profits, what would this incentive be with a European-style waterfall whole-of-fund approach, and what would it be if the incentive is paid on an American-style waterfall deal-by-deal basis (assuming no clawback applies)?

Solution:

In aggregate, the fund lost money (+\$9 million – \$10 million = –\$1 million), so with a European-style whole-of-fund waterfall and assuming the time period for the gain and the loss are the same, there is no incentive fee. With an American-style waterfall, the GP could still earn 20% × \$9 million = \$1.8 million on the yacht company, thereby further compounding the loss to the ultimate investor to –\$2.8 million net of fees.

	Aggregate	Yacht Company	Casino Venture
Investment	\$32 m	\$15 m	\$17 m
Profit/Loss	-\$1 m	\$9 m	-\$10 m
Incentive by Deal		20% × \$9m = \$1.8m	\$0 m
Total Incentive	\$0 m	\$1.8 m	

If the gain and loss in this example transpired sequentially over different years, perhaps with the yacht company gain occurring first and then the Casino venture loss later on, the outcome for a European-style waterfall would typically result in an initial accrued incentive fee for the yacht manufacturer gain, but if there is a clawback provision in place, then there would be a clawback of that fee for the investor in the subsequent year when the casino venture loss is eventually realized, still resulting in no overall incentive fee. Waterfall language and clawback provisions on fees are very important to study and understand in offering memorandums, and these terms can vary widely.

10.2 Custom Fee Arrangements

Although "2 and 20" and "1 and 10" are commonly quoted fee structures for hedge funds and funds of funds, respectively, many variations exist.

1 *Fees based on liquidity terms and asset size:* Hedge funds may charge different rates depending on the liquidity terms that an investor is willing to accept (longer lockups are generally associated with lower fees), and hedge fund managers may discount their fees for larger investors or for placement agents who introduced these investors. Different investors in the same fund may well face different fee structures. Hedge fund managers negotiate terms, including fees and notice and lockup periods, with individual investors via side letters, which are special amendments to a standard offering memorandum's terms and conditions. For perspective, management fees for large LPs could range from 0.5% to 1.5%, with incentive fees reduced to 10%–15%, depending on the mandate. Such reductions can be meaningful in terms of net realized returns. However, some smaller hedge funds with strong performance (and capacity constraints) are able to maintain higher fees and may even turn down business from larger investors rather than agree to a lower fee.

2 *Founders' shares*: As a way to entice early participation in start-up and emerging hedge funds, managers have increasingly offered incentives known as founders' class shares. Founders' shares entitle investors to a lower fee structure, such as 1.5 and 10 rather than 2 and 20, and are typically available to be applied only to the first \$100 million in assets, although cutoff thresholds vary. Another option is to reduce the fees for early founders' share investors once the fund achieves a

critical mass of assets or performance targets. Both paths act as an incentive to spur investors to make faster investment commitments than might otherwise be the case.

3 *"Either/or" fees*: As a pushback against high hedge fund fees, institutional investors, such as the Teacher Retirement System of Texas, have recently requested a new fee model that some hedge fund managers have started to accept in return for substantive allocations. Managers agree *either* to charge a 1% management fee (simply to cover expenses during down years) *or* to receive a 30% incentive fee above a mutually agreed-on annual hurdle (to incentivize and reward managers during up years), whichever is greater. If a manager were to go without profits for a year or two, the 1% management fee becomes effectively an advance against an eventual 30% incentive-fee year (thereby reducing that future-year incentive fee by the prior years' advanced management fees. Although a far cry from the traditional 2 and 20, such novel fee structures, which are designed to reward performance and the delivery of true alpha above a benchmark, are likely to become even more in demand in the institutional hedge fund industry. Hedge funds will likely also endeavor to charge high-net-worth investors (with smaller commitment sizes) more traditional fees.

The following example demonstrates fee structures and their effect on the resulting returns to investors.

EXAMPLE 5

Fee and Return Calculations

AWJ Capital is a hedge fund with $100 million of initial investment capital. It charges a 2% management fee based on year-end AUM and a 20% incentive fee. In its first year, AWJ Capital has a 30% return. Assume management fees are calculated using end-of-period valuation.

1 What are the fees earned by AWJ if the incentive and management fees are calculated independently? What is an investor's effective return given this fee structure?

2 What are the fees earned by AWJ assuming that the incentive fee is calculated from the return net of the management fee? What is an investor's net return given this fee structure?

3 If the fee structure specifies a hurdle rate of 5% and the incentive fee is based on returns in excess of the hurdle rate, what are the fees earned by AWJ assuming the performance fee is calculated net of the management fee? What is an investor's net return given this fee structure?

4 In the second year, the fund value declines to $110 million. The fee structure is as specified for Question 1 but also includes the use of a high-water mark (computed net of fees). What are the fees earned by AWJ in the second year? What is an investor's net return for the second year given this fee structure?

5 In the third year, the fund value increases to $128 million. The fee structure is as specified in Questions 1 and 4. What are the fees earned by AWJ in the third year? What is an investor's net return for the third year given this fee structure?

Solution to 1:

AWJ fees

$130 million × 2% = $2.6 million management fee.

($130 – $100) million × 20% = $6 million incentive fee.

Total fees to AWJ Capital = $8.6 million.

Investor return: ($130 – $100 – $8.6) million/$100 million = 21.40%.

Solution to 2:

$130 million × 2% = $2.6 million management fee.

($130 – $100 – $2.6) million × 20% = $5.48 million incentive fee.

Total fees to AWJ Capital = $8.08 million.

Investor return: ($130 – $100 – $8.08) million/$100 million = 21.92%.

Solution to 3:

$130 million × 2% = $2.6 million management fee.

($130 – $100 – $5 – $2.6) million × 20% = $4.48 million incentive fee.

Total fees to AWJ Capital = $7.08 million.

Investor return: ($130 – $100 – $7.08) million/$100 million = 22.92%.

Solution to 4:

$110 million × 2% = $2.2 million management fee.

No incentive fee because the fund has declined in value.

Total fees to AWJ Capital = $2.2 million.

Investor return: ($110 – $2.2 – $121.4) million/$121.4 million = –11.20%.

The beginning capital position in the second year for the investors is ($130 – $8.6) million = $121.4 million. The ending capital position at the end of the second year is ($110 – $2.2) million = $107.8 million.

Solution to 5:

$128 million × 2% = $2.56 million management fee.

($128 – $121.4) million × 20% = $1.32 million incentive fee.

The $121.4 million represents the high-water mark established at the end of Year 1.

Total fees to AWJ Capital = $3.88 million.

Investor return: ($128 – $3.88 – $107.8) million/$107.8 million = 15.14%.

The ending capital position at the end of Year 3 is $124.12 million. This amount is the new high-water mark.

As the previous example illustrates, the return to an LP investor in a fund may differ significantly from the quoted return for the fund as a whole, which generally reflects the return that a “Day 1” investor who made no capital movements would have earned. Hedge fund databases and indexes generally report performance net of fees. If fee structures vary, however, the actual net-of-fees returns earned by various investors often may vary from the one included in a given database or index.

The multi-layered fee structure of funds of hedge funds has the effect of further diluting returns to the investor, but as discussed, this disadvantage may be balanced by positive features, such as access to a diversified portfolio and to hedge funds that may otherwise be closed to direct investments, as well as expertise in due diligence in hedge fund selection. There is thus both added cost and added value.

Generally, over time many funds of funds have earned a reputation for being "fast" money because their managers tend to be the first to redeem their investment when a hedge fund performs poorly. They may also have negotiated more favorable redemption terms—a shorter lockup or notice period, for example. Because of the overall compression of hedge fund returns and issues of excessive fee layering, many fund-of-funds managers have been pressured to drop their incentive fees and simply charge a flat management fee.

KNOWLEDGE CHECK

Fill in the blank: Fund offering documents typically offer terms that include a ___________, whereby incentive fees will accrue, and subsequently be paid, only on new earnings above and beyond the recoupment of any prior losses.

Solution:

Fund offering documents typically offer terms that include a high-water mark, whereby incentive fees will accrue, and subsequently be paid, only on new earnings above and beyond the recoupment of any prior losses.

KNOWLEDGE CHECK

True or false: Advantages of funds-of-hedge funds include: due diligence in selecting individual hedge funds, access to hedge funds that may be closed to direct investments, and dilution of returns to the investor.

A True, statement to support this option on why it's true.

B False, statement to support this option on why it's false.

Solution:

B is correct; the statement is false. Although these three attributes are indeed true of funds of hedge funds, the dilution of returns to the investor is a disadvantage for the investor, not an advantage. The "due diligence" and "access" attributes are advantages.

EXAMPLE 6

Comparison of Returns: Investment Directly into a Hedge Fund or through a Fund of Hedge Funds

An investor is contemplating investing €100 million in either the ABC Hedge Fund (ABC HF) or the XYZ Fund of Funds (XYZ FOF). XYZ FOF has a "1 and 10" fee structure and invests 10% of its AUM in ABC HF. ABC HF has a standard "2 and 20" fee structure with no hurdle rate. Management fees are calculated on an annual basis on AUM at the beginning of the year. For simplicity, assume that management fees and incentive fees are calculated independently. ABC HF has a 20% return for the year before management and incentive fees.

1 Calculate the return to the investor from investing directly in ABC HF.
2 Calculate the return to the investor from investing in XYZ FOF. Assume that the other investments in the XYZ FOF portfolio generate the same return before management fees as those of ABC HF and that XYZ FOF has the same fee structure as ABC HF.
3 Why would the investor choose to invest in a fund of funds instead of a hedge fund given the effect of the "double fee" demonstrated in the answers to Questions 1 and 2?

Solution to 1:

ABC HF has a profit before fees on a €100 million investment of €20 million (= €100 million × 20%). The management fee is €2 million (= €100 million × 2%), and the incentive fee is €4 million (= €20 million × 20%). The return to the investor is 14% [= (20 – 2 – 4)/100].

Solution to 2:

XYZ FOF earns a 14% return or €14 million profit after fees on €100 million invested with hedge funds. XYZ FOF charges the investor a management fee of €1 million (= €100 million × 1%) and an incentive fee of €1.4 million (= €14 million × 10%). The return to the investor is 11.6% [= (14 – 1 – 1.4)/100].

Solution to 3:

This scenario assumes that returns are the same for all underlying hedge funds. In practice, this result will not likely be the case, and XYZ FOF may provide due diligence expertise and potentially valuable diversification.

10.3 Alignment of Interests and Survivorship Bias

The alternative asset business is attractive to portfolio managers because of how significant the fees can be if the fund performs well and AUM are significant. But as discussed previously, high fees destroy value and reduce the attractiveness of alternative investing. If a hedge fund manager can stay in business for just four to five years with acceptable returns, incentive fee allocations to the general partner can be substantive.

Comparatively, the commitment of a private equity or real estate manager to stay in business for 6–10 years (often with offering memorandum language that first attempts to return all capital to investors, then attempts in many instances to pay a preferred minimum return to investors, and then only belatedly allows a manager to crystalize, or realize, its own incentive fees) is arguably a more aligned overall incentive structure. However, the overall time commitment for the investor is obviously of a much longer duration. The investor runs a larger risk of being stuck in something that might be a disappointment, as opposed to an ability to simply move on, as with hedge funds. Landing a private equity commitment basically guarantees a manager a very attractive management fee runway for an extended period of time.

So, there remain these types of trade-offs that still make the alternative investing world overall quite lucrative regardless of the specific fund or fee format. Because of this generally lucrative incentive structure—particularly for hedge funds and their ability to crystalize incentive fees (even on gains that may still be unrealized) on an annualized basis—many new hedge funds launched in the late 1990s and early 2000s. Not all hedge funds, however, remain in business for very long. One study suggests that more than a quarter of all hedge funds fail within the first three years because of performance problems that result in investor defections and the ultimate failure to generate sufficient revenue to cover the fund's operating costs. This is one reason survivorship bias is a major problem with hedge fund indexes. *Backfill bias* is another problem: Certain surviving hedge funds may be added to databases and various hedge fund indexes only after they are initially successful and start to report their returns. Because of survivorship and backfill biases, hedge fund indexes may not reflect actual hedge fund performance but, rather, reflect only the performance of hedge funds that are performing well and thus "survived."

EXAMPLE 7

1 Clawbacks Due to Return Timing Differences

The Granite Rock Fund makes investments in leveraged-buyout Company A and start-up Company B, each for $10 million. One year later, the leveraged-buyout company returns a $14 million profit, and two years later, the start-up company turns into a complete bust, deemed to be worth zero.

If the GP's carried interest incentive fee is 20% of aggregate profits and there is a clawback provision, how much carried interest will the GP initially accrue and ultimately receive?

Solution:

From leveraged-buyout Company A, the GP would initially accrue a 20% of $14 million profit at the end of the first year, equal to $2.8 million. Typically, this amount would be held in escrow for the benefit of the GP but not actually paid.

But then the GP loses $10 million of the initial $14 million gain, so the aggregate whole-of-fund gain at the end of the second year would be only $4 million; this amount times 20% would result in only an $800,000 incentive fee. The general partner would then have to return $2 million of the previously accrued incentive fees to the capital accounts of LP investors because of the clawback provision.

2 Soft and Hard Hurdles

A real estate investment fund has a $100 million initial drawdown structure in its first year and fully draws this capital to purchase a property. The fund has a soft hurdle preferred return to investors of 8% per annum and an 80%/20% carried interest incentive split thereafter. At the end of year two, the property is sold for a total of $160 million.

What are the correct distributions to the LPs and to the GP? And how would these have been different if the real estate investment fund had a hard hurdle of 8% per annum instead of a soft hurdle?

Solution:

One needs to construct a waterfall of cash flows.

First, the LPs would be due their $100 million initial investment.

Then, they would be due $16 million (8% preferred return on initial capital for two years).

The soft hurdle has been met, and the GP is ultimately due 20% of $60 million, or $12 million, which would be paid to the GP next as a catch-up to the achieved hurdle return.

The residual amount would be $160 million – $100 million – $16 million – $12 million = $32 million. This amount would then be split 80% to the LPs and 20% to the GP, or $25.6 million and $6.4 million, respectively.

So, the total payout with a soft annual hurdle of 8% of the $160 million would end up with the following waterfall:

	LP	GP
Return of Capital	\$100 m	
8% Prfd per Annum	\$16 m	
GP Catch-Up 20%		\$12 m
80%/20% Split	\$25.6 m	\$6.4 m
Total Payout	\$141.6 m	18.4 m

If the fund instead had a hard hurdle rate, only the amount above the \$100 return of capital and \$16 million preferred return would be subject to the 20% carried interest incentive to the GP: 20% × \$44 million = \$8.8 million, quite a bit less than the carried interest payment with the soft hurdle. The LPs would be due the balance of \$35.2 million (\$44 million – \$8.8 million incentive). This would result in the following total payout:

	LP	GP
Return of Capital	\$100 m	
8% Prfd per Annum	\$16 m	
80%/20% Split Above Hurdle	\$35.2 m	\$8.8 m
Total Payout	\$151.2 m	8.8 m

PRACTICE QUESTIONS FOR LOS J

Question 1

United Capital is a hedge fund with \$250 million of initial capital. United charges a 2% management fee based on assets under management at year end and a 20% incentive fee based on returns in excess of an 8% hurdle rate. In its first year, United appreciates 16%. Assume management fees are calculated using end-of-period valuation. The investor's net return assuming the performance fee is calculated net of the management fee is *closest* to: (2020 Q26)

A 11.58%.

B 12.54%.

C 12.80%.

Solution:

B is correct. The net investor return is 12.54%, calculated as follows:

End-of-year capital = \$250 million × 1.16 = \$290 million.

Management fee = \$290 million × 2% = \$5.8 million.

Hurdle amount = 8% of \$250 million = \$20 million.

Incentive fee = (\$290 – \$250 – \$20 – \$5.8) million × 20% = \$2.84 million.

Total fees to United Capital = (\$5.8 + \$2.84) million = \$8.64 million.

Investor net return: (\$290 – \$250 – \$8.64)/\$250 = 12.54%.

Question 2

Capricorn Fund of Funds invests GBP100 million in each of Alpha Hedge Fund and ABC Hedge Fund. Capricorn Fund of Funds has a "1 and 10" fee structure. Management fees and incentive fees are calculated independently at the end of each year. After one year, net of their respective management and incentive fees, Capricorn's investment in Alpha is valued at GBP80 million and Capricorn's investment in ABC is valued at GBP140 million. The annual return to an investor in Capricorn Fund of Funds, net of fees assessed at the fund-of-funds level, is *closest* to: (2020 Q27)

A 7.9%.

B 8.0%.

C 8.1%.

Solution:

A is correct because the net investor return is 7.9%, calculated as follows:

First, note that "1 and 10" refers to a 1% management fee and a 10% incentive fee.

End-of-year capital = GBP140 million + GBP80 million = GBP220 million.

Management fee = GBP220 million × 1% = GBP2.2 million.

Incentive fee = (GBP220 − GBP200) million × 10% = GBP2 million.

Total fees to Capricorn = (GBP2.2 + GBP2) million = GBP4.2 million.

Investor net return: (GBP220 − GBP200 − GBP4.2)/GBP200 = 7.9%.

Question 3

The following information applies to Rotunda Advisers, a hedge fund:

- $288 million in AUM as of prior year end
- 2% management fee (based on year-end AUM)
- 20% incentive fee calculated:
  - net of management fee
  - using a 5% soft hurdle rate
  - using a high-water mark (high-water mark is $357 million)
- Current-year fund gross return is 25%.

The total fee earned by Rotunda in the current year is *closest* to: (2020 Q28)

A $7.20 million.

B $20.16 million.

C $21.60 million.

Solution:

A is correct. Although the gross return of Rotunda results in a $360 million gross NAV, the deduction of the $7.2 million incentive fee brings NAV to $352.8 million, which is below the prior high-water mark. Rotunda earns a management fee of $7.20 million but does not earn an incentive fee because the year-end fund value net of management fee does not exceed the prior high-water mark of $357 million. Since Rotunda is still also below the prior-year high-water mark, the hurdle rate of return is also basically irrelevant in this fee calculation.

The specifics of this calculation are as follows:

End-of-year AUM = Prior year-end AUM × (1 + Fund return) = $288 million × 1.25 = $360 million.

$360 million × 2% = $7.20 million management fee.

$360 million − $7.2 million = $352.8 million AUM net of management fee.

The year-end AUM net of fees do not exceed the $357 million high-water mark. Therefore, no incentive fee is earned.

Question 4

A hedge fund has the following fee structure:

Annual management fee based on year-end AUM	2%
Incentive fee	20%
Hurdle rate before incentive fee collection starts	4%
Current high-water mark	$610 million

The fund has a value of $583.1 million at the beginning of the year. After one year, it has a value of $642 million before fees. The net percentage return to an investor for this year is *closest* to: (2020 Q29)

A 6.72%.

B 6.80%.

C 7.64%.

Solution:

C is correct. The management fee for the year is

$642 × 0.02 = $12.84 million.

Because the ending gross value of the fund of $642 million exceeds the high-water mark of $610 million, the hedge fund can collect an incentive fee on gains above this high-water mark but net of the hurdle rate of return. The incentive fee calculation becomes

{$642 − [$610 × (1 + 0.04)]} × 0.20 = $1.52 million.

The net return to the investor for the year is

[($642 − $12.84 − $1.52)/$583.1] − 1 ≈ 0.07638 ≈ 7.64%.

SUMMARY

This reading provides a comprehensive introduction to alternative investments. Some key points of the reading are as follows:

- Alternative investments are supplemental strategies to traditional long-only positions in stocks, bonds, and cash. Alternative investments include investments in five main categories: hedge funds, private capital, natural resources, real estate, and infrastructure.
- Alternative investment strategies are typically active, return-seeking strategies that also often have risk characteristics different from those of traditional long-only investments.

- Characteristics common to many alternative investments, when compared with traditional investments, include the following: lower liquidity, less regulation, lower transparency, higher fees, and limited and potentially problematic historical risk and return data.
- Alternative investments often have complex legal and tax considerations and may be highly leveraged.
- Alternative investments are attractive to investors because of the potential for portfolio diversification resulting in a higher risk-adjusted return for the portfolio.
- Investors can access alternative invests in three ways:
  - Fund investment (such as a in a PE fund)
  - Direct investment into a company or project (such as infrastructure or real estate)
  - Co-investment into a portfolio company of a fund
- Investors conduct due diligence prior to investing in alternative investments. The due diligence approach depends on the investment method (direct, co-investing, or fund investing).
- Operational, financial, counterparty, and liquidity risks may be key considerations for those investing in alternative investments. These risks can be analyzed during the due diligence process. It is critical to perform fund due diligence to assess whether (a) the manager can effectively pursue the proposed investment strategy; (b) the appropriate organizational structure and policies for managing investments, operations, risk, and compliance are in place; and (c) the fund terms appear reasonable.
- Many alternative investments, such as hedge and private equity funds, use a partnership structure with a general partner that manages the business and limited partners (investors) who own fractional interests in the partnership.
- The general partner typically receives a management fee based on assets under management or committed capital (the former is common to hedge funds, and the latter is common to private equity funds) and an incentive fee based on realized profits.
- Hurdle rates, high-water marks, lockup and notice periods, and clawback provisions are often specified in the LPA.
- The fee structure affects the returns to investors (limited partners), with a waterfall representing the distribution method under which allocations are made to LPs and GPs. Waterfalls can be on a whole-of-fund basis (European) or deal-by-deal basis (American).
- Hedge funds are typically classified by strategy. One such classification includes four broad categories of strategies: equity hedge (e.g., market neutral), event driven (e.g., merger arbitrage), relative value (e.g., convertible bond arbitrage), macro and CTA strategies (e.g., commodity trading advisers).
- Funds-of-hedge-funds are funds that create a diversified portfolio of hedge funds. These vehicles are attractive to smaller investors that don't have the resources to select individual hedge funds and build a portfolio of them.
- Private Capital is a broad term for funding provided to companies that is sourced from neither the public equity nor debt markets. Capital that is provided in the form of equity investments is called private equity, whereas capital that is provided as a loan or other form of debt is called private debt.

- Private equity refers to investment in privately owned companies or in public companies with the intent to take them private. Key private equity investment strategies include leveraged buyouts (e.g., MBOs and MBIs) and venture capital. Primary exit strategies include trade sale, IPO, and recapitalization.
- Private debt refers to various forms of debt provided by investors to private entities. Key private debt strategies include direct lending, mezzanine debt, and venture debt. Private debt also includes specialized strategies, such as CLOs, unitranche debt, real estate debt, and infrastructure debt.
- Natural resources include commodities (hard and soft), agricultural land (farmland), and timberland.
- Commodity investments may involve investing in actual physical commodities or in producers of commodities, but more typically, these investments are made using commodity derivatives (futures or swaps). One can also invest in commodities via a CTA (see hedge funds)
- Returns to commodity investing are based on changes in price and do not include an income stream, such as dividends, interest, or rent (apart from income earned on the collateral). However, timberland offers an income stream based on the sale of trees, wood, and other products. Timberland can be thought of as both a factory and a warehouse. Plus, timberland is a sustainable investment that mitigates climate-related risks.
- Farmland, like timberland, has an income component related to harvest quantities and agricultural commodity prices. However, farmland doesn't have the production flexibility of timberland, because farm products must be harvested when ripe.
- Real estate includes two major sectors: residential and commercial. Residential real estate is the largest sector, making up some 75% of the market globally. Commercial real estate primarily includes office buildings, shopping centers, and warehouses. Real estate property has some unique features compared with other asset classes, including heterogeneity (no two properties are identical) and fixed location.
- Real estate investments can be direct or indirect, in the public market (e.g., REITs) or private transactions, and in equity or debt.
- The assets underlying infrastructure investments are real, capital intensive, and long lived. These assets are intended for public use, and they provide essential services. Examples include airports, health care facilities, and power plants. Funding is often done on a public–private partnership basis.
- Social infrastructure assets are directed toward human activities and include such assets as educational, health care, social housing, and correctional facilities, with the focus on providing, operating, and maintaining the asset infrastructure.
- Infrastructure investments may also be categorized by the underlying asset's stage of development. Investing in infrastructure assets *that are to be constructed* is generally referred to as greenfield investment. Investing in *existing* infrastructure assets may be referred to as brownfield investment.
- Conducting performance appraisal on alternative investments can be challenging because these investments are often characterized by asymmetric risk–return profiles, limited portfolio transparency, illiquidity, product complexity, and complex fee structures.

- Traditional risk and return measures (such as mean return, standard deviation of returns, and beta) may provide an inadequate picture of alternative investments' risk and return characteristics. Moreover, these measures may be unreliable or not representative of specific investments.
- A variety of ratios can be calculated in order to review the performance of alternative investments, including the Sharpe ratio, Sortino ratio, Treynor ratio, Calmar ratio, and MAR ratio. In addition, batting average and slugging percentage can also be used. The IRR calculation is often used to evaluate private equity investments, and the cap rate is often used to evaluate real estate investments.
- Redemption rules and lockup periods can bring special challenges to performance appraisal of alternative investments.
- When comparing the performance of alternative investments versus an index, the analyst must be aware that indexes for alternative investments may be subject to a variety of biases, including survivorship and backfill biases.
- Analysts need to be aware of any custom fee arrangements in place that will affect the calculation of fees and performance. These can include such arrangements as fee discounts based on custom liquidity terms or significant asset size; special share classes, such as "founders' shares"; and a departure from the typical management fee + performance fee structure in favor of "either/or" fees

Portfolio Management

STUDY SESSION

Study Session 17	Portfolio Management (1)
Study Session 18	Portfolio Management (2)

TOPIC LEVEL LEARNING OUTCOME

The candidate should be able to explain and demonstrate the use of fundamentals of portfolio and risk management, including return and risk measurement, and portfolio planning and construction.

PORTFOLIO MANAGEMENT
STUDY SESSION

17

Portfolio Management (1)

This study session introduces the concept of a portfolio approach to investments. The needs of individual and institutional investors are each examined, along with the range of available investment solutions. The three main steps in the portfolio management process (planning, execution, and feedback) are outlined. Common measures of portfolio risk and return and the introduction of modern portfolio theory—a quantitative framework for asset pricing and portfolio selection—then follow.

READING ASSIGNMENTS

Reading 48	Portfolio Management: An Overview by Owen M. Concannon, CFA, Robert M. Conroy, DBA, CFA, Alistair Byrne, PhD, CFA, and Vahan Janjigian, PhD, CFA
Reading 49	Portfolio Risk and Return: Part I by Vijay Singal, PhD, CFA
Reading 50	Portfolio Risk and Return: Part II by Vijay Singal, PhD, CFA

READING

48

Portfolio Management: An Overview

by Owen M. Concannon, CFA, Robert M. Conroy, DBA, CFA, Alistair Byrne, PhD, CFA, and Vahan Janjigian, PhD, CFA

Owen M. Concannon, CFA, is at Neuberger Berman (USA). Robert M. Conroy, DBA, CFA, is at the Darden School of Business, University of Virginia (USA). Alistair Byrne, PhD, CFA, is at State Street Global Advisors (United Kingdom). Vahan Janjigian, PhD, CFA, is at Greenwich Wealth Management, LLC (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** describe the portfolio approach to investing;
☐	**b.** describe the steps in the portfolio management process;
☐	**c.** describe types of investors and distinctive characteristics and needs of each;
☐	**d.** describe defined contribution and defined benefit pension plans;
☐	**e.** describe aspects of the asset management industry;
☐	**f.** describe mutual funds and compare them with other pooled investment products.

INTRODUCTION

1

This reading provides an overview of portfolio management and the asset management industry, including types of investors and investment plans and products. A portfolio approach is important to investors in achieving their financial objectives. We outline the steps in the portfolio management process in managing a client's investment portfolio. We next compare the financial needs of different types of investors: individual and institutional investors. We then describe both defined contribution and defined benefit pension plans. The asset management[1] industry, which serves as a critical link between providers and seekers of investment capital around the world, is broadly discussed. Finally, we describe mutual funds and other types of pooled investment products offered by asset managers.

1 Note that both "investment management" and "asset management" are commonly used throughout the CFA Program curriculum. The terms are often used interchangeably in practice.

2 PORTFOLIO PERSPECTIVE: DIVERSIFICATION AND RISK REDUCTION

a describe the portfolio approach to investing;

One of the biggest challenges faced by individuals and institutions is to decide how to invest for future needs. For individuals, the goal might be to fund retirement needs. For such institutions as insurance companies, the goal is to fund future liabilities in the form of insurance claims, whereas endowments seek to provide income to meet the ongoing needs of such institutions as universities. Regardless of the ultimate goal, all face the same set of challenges that extend beyond just the choice of what asset classes to invest in. They ultimately center on formulating basic principles that determine how to think about investing. One important question is: Should we invest in individual securities, evaluating each in isolation, or should we take a portfolio approach? By "portfolio approach," we mean evaluating individual securities in relation to their contribution to the investment characteristics of the whole portfolio. In the following section, we illustrate a number of reasons why a diversified portfolio perspective is important.

2.1 Historical Example of Portfolio Diversification: Avoiding Disaster

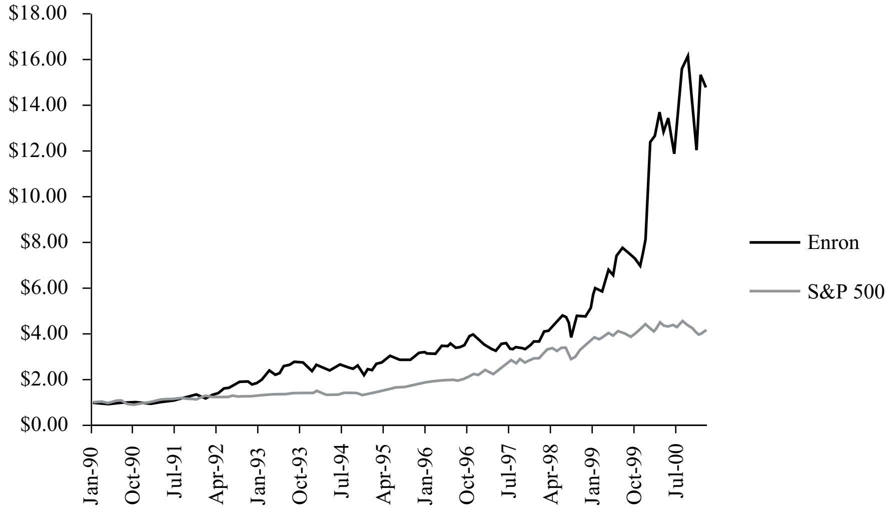
Portfolio diversification helps investors avoid disastrous investment outcomes. This benefit is most convincingly illustrated by examining what may happen when individuals have *not* diversified.

We are usually not able to observe how individuals manage their personal investments. However, in the case of US 401(k) individual retirement portfolios,[2] it is possible to see the results of individuals' investment decisions. When we examine their retirement portfolios, we find that some individual participants make sub-optimal investment decisions.

During the 1990s, Enron Corporation was one of the most admired corporations in the United States. A position in Enron shares returned over 27 percent per year from 1990 to September 2000, compared to 13 percent for the S&P 500 Index for the same time period.

2 In the United States, 401(k) plans are employer-sponsored individual retirement savings plans. They allow individuals to save a portion of their current income and defer taxation until the time when the savings and earnings are withdrawn. In some cases, the sponsoring firm will also make matching contributions in the form of cash or shares. Individuals within certain limits have control of the invested funds and consequently can express their preferences as to which assets to invest in.

Exhibit 1 Value of US$1 Invested from January 1990 to September 2000 Enron vs. S&P 500 Composite Index (01/01/1990 = US$1.00)



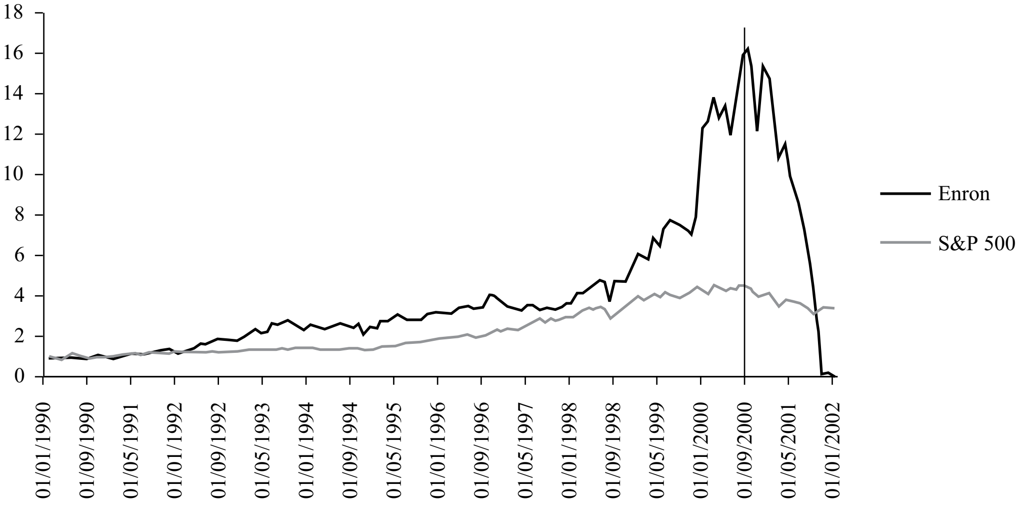
Source: Thomson Reuters Datastream.

During this time period, thousands of Enron employees participated in the company's 401(k) retirement plan. The plan allowed employees to set aside some of their earnings in a tax-deferred account. Enron participated by matching the employees' contributions. Enron made the match by depositing required amounts in the form of Enron shares. Enron restricted the sale of its contributed shares until an employee turned 50 years old. In January 2001, the employees' 401(k) retirement accounts were valued at over US$2 billion, of which US$1.3 billion (or 62 percent) was in Enron shares. Although Enron restricted the sale of shares it contributed, less than US$150 million of the total of US$1.3 billion in shares had this restriction. The implication was that Enron employees continued to hold large amounts of Enron shares even though they were free to sell them and invest the proceeds in other assets.

A typical individual was Roger Bruce,[3] a 67-year-old Enron retiree who held all of his US$2 million in retirement funds in Enron shares. Between January 2001 and January 2002, Enron's share price fell from about US$90 per share to zero.

3 Singletary (2001).

Exhibit 2 Value of US$1 Invested from January 1990 to January 2002 Enron vs. S&P 500 Composite Index (1/1/1990 = US$1.00)



Source: Thomson Reuters Datastream.

Employees and retirees who had invested all or most of their retirement savings in Enron shares, just like Mr. Bruce, experienced financial ruin. The hard lesson that the Enron employees learned from this experience was to "not put all your eggs in one basket."[4] Unfortunately, the typical Enron employee did have most of his or her eggs in one basket. Most employees' wages and financial assets were dependent on Enron's continued viability; hence, any financial distress on Enron would have a material impact on an employee's financial health. The bankruptcy of Enron resulted in the closing of its operations, the dismissal of thousands of employees, and its shares becoming worthless. Hence, the failure of Enron was disastrous to the typical Enron employee.

Enron employees were not the only ones to be victims of over-investment in a single company's shares. In the defined contribution retirement plans at Owens Corning, Northern Telecom, Corning, and ADC Telecommunications, employees all held more than 25 percent of their assets in the company's shares during a time (March 2000 to December 2001) in which the share prices in these companies fell by almost 90 percent. The good news in this story is that the employees participating in employer-matched 401(k) plans since 2001 have significantly reduced their holdings of their employers' shares.

Thus, by taking a diversified portfolio approach, investors can spread away some of the risk. Rational investors are concerned about the risk–return trade-off of their investments. The portfolio approach provides investors with a way to reduce the risk associated with their wealth without necessarily decreasing their expected rate of return.

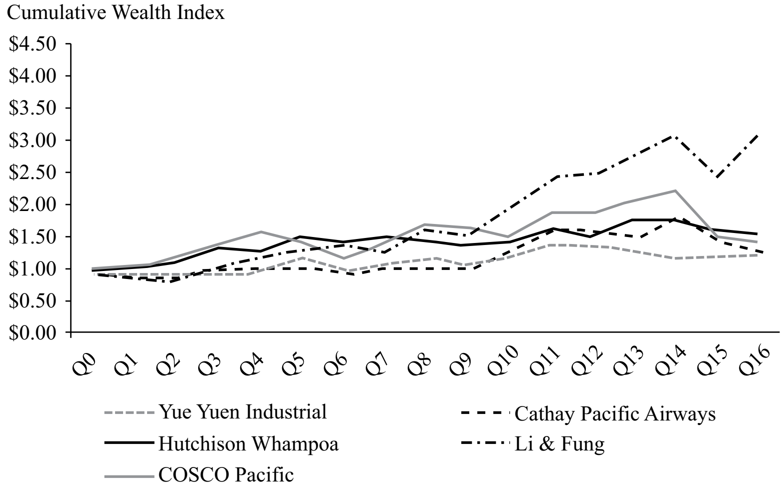
2.2 Portfolios: Reduce Risk

In addition to avoiding a potential disaster associated with over investing in a single security, portfolios also generally offer equivalent expected returns with lower overall **volatility** of returns—as represented by a measure such as standard deviation. Consider

4 This expression, which most likely originated in England in the 1700s, has a timeless sense of wisdom.

this simple example: Suppose you wish to make an investment in companies listed on the Hong Kong Stock Exchange (HKSE) and you start with a sample of five companies.[5] The cumulative returns for 16 fiscal quarters are shown in Exhibit 3.

Exhibit 3 Cumulative Wealth Index of Sample of Shares Listed on HKSE (initial amount= US$1.00)



Source: Thomson Reuters Datastream.

The individual quarterly returns for each of the five shares are shown in Exhibit 4. The annualized means and annualized standard deviations for each are also shown.[6]

Exhibit 4 Quarterly Returns (in Percent) for Sample of HKSE Listed Shares over 16 Fiscal Quarters

	Yue Yuen Industrial	Cathay Pacific Airways	Hutchison Whampoa	Li & Fung	COSCO Pacific	Equally Weighted Portfolio
Q1	−11.1%	−2.3%	0.6%	−13.2%	−1.1%	−5.4%
Q2	−0.5	−5.4	10.8	1.7	21.0	5.5
Q3	5.7	6.8	19.1	13.8	15.5	12.2
Q4	5.3	4.6	−2.1	16.9	12.4	7.4
Q5	17.2	2.4	12.6	14.5	−7.9	7.8
Q6	−17.6	−10.4	−0.9	4.4	−16.7	−8.2
Q7	12.6	7.4	4.2	−10.9	15.4	5.7
Q8	7.5	−0.4	−3.6	29.2	21.9	10.9
Q9	−7.9	1.3	−5.1	−2.0	−1.6	−3.1
Q10	8.2	27.5	0.1	26.0	−10.1	10.3
Q11	18.3	24.3	16.5	22.8	25.7	21.5
Q12	0.1	−2.6	−6.7	−0.4	0.3	−1.8

(continued)

5 A sample of five companies from a similar industry group was arbitrarily selected for illustration purposes.
6 Mean quarterly returns are annualized by multiplying the quarterly mean by 4. Quarterly standard deviations are annualized by taking the quarterly standard deviation and multiplying it by 2.

Exhibit 4 (Continued)

	Yue Yuen Industrial	Cathay Pacific Airways	Hutchison Whampoa	Li & Fung	COSCO Pacific	Equally Weighted Portfolio
Q13	−6.2	−4.2	16.7	11.9	11.1	5.8
Q14	−8.0	17.9	−1.8	12.4	8.4	5.8
Q15	3.5	−20.1	−8.5	−20.3	−31.5	−15.4
Q16	2.1	−11.8	−2.6	24.2	−6.1	1.2
Mean annual return	7.3%	8.7%	12.3%	32.8%	14.2%	15.1%
Annual standard deviation	20.2%	25.4%	18.1%	29.5%	31.3%	17.9%
Diversification ratio						71.9%

Source: Thomson Reuters Datastream.

Suppose you want to invest in one of these five securities next year. There is a wide variety of risk–return trade-offs for the five shares selected. If you believe that the future will replicate the past, then choosing Li & Fung would be a good choice. For the prior four years, Li & Fung provided the best trade-off between return and risk. In other words, it provided the most return per unit of risk. However, if there is no reason to believe that the future will replicate the past, it is more likely that the risk and return on the one security selected will be more like selecting one randomly. When we randomly selected one security each quarter, we found an average annualized return of 15.1 percent and an average annualized standard deviation of 24.9 percent, which would now become your expected return and standard deviation, respectively.

Alternatively, you could invest in an equally weighted portfolio of the five shares, which means that you would invest the same dollar amount in each security for each quarter. The quarterly returns on the equally weighted portfolio are just the average of the returns of the individual shares. As reported in Exhibit 4, the equally weighted portfolio has an average return of 15.1 percent and a standard deviation of 17.9 percent. As expected, the equally weighted portfolio's return is the same as the return on the randomly selected security. However, the same does not hold true for the portfolio standard deviation. That is, the standard deviation of an equally weighted portfolio is not simply the average of the standard deviations of the individual shares. In a more advanced reading we will demonstrate in greater mathematical detail how such a portfolio offers a lower standard deviation of return than the average of its individual components due to the correlations or interactions between the individual securities.

Because the mean return is the same, a simple measure of the value of diversification is calculated as the ratio of the standard deviation of the equally weighted portfolio to the standard deviation of the randomly selected security. This ratio may be referred to as the **diversification ratio**. In this case, the equally weighted portfolio's standard deviation is approximately 72 percent of the average standard deviation of the 5 stocks (24.9%). The diversification ratio of the portfolio's standard deviation to the individual asset's standard deviation measures the risk reduction benefits of a simple portfolio construction method, equal weighting. Even though the companies were chosen from a similar industry grouping, we see significant risk reduction. An even greater portfolio effect (i.e., lower diversification ratio) could have been realized if we had chosen companies from completely different industries.

This example illustrates one of the critical ideas about portfolios: Portfolios affect risk more than returns. In the prior section portfolios helped avoid the effects of downside risk associated with investing in a single company's shares. In this section we extended the notion of risk reduction through portfolios to illustrate why individuals and institutions should hold portfolios.

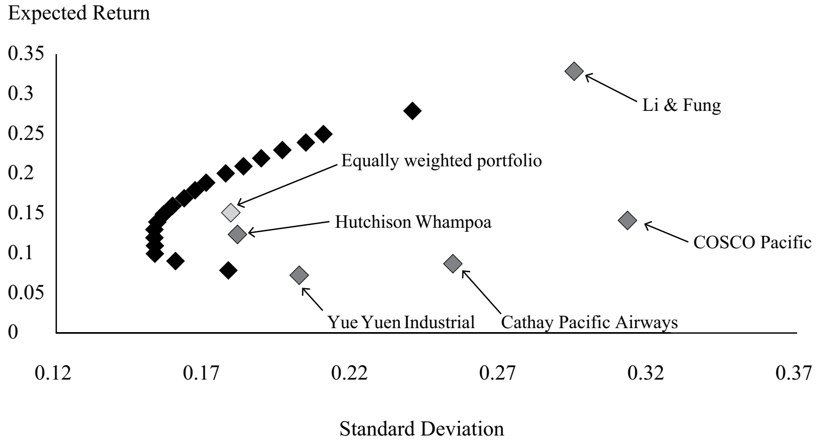
PORTFOLIO PERSPECTIVE: RISK-RETURN TRADE-OFF, DOWNSIDE PROTECTION, MODERN PORTFOLIO THEORY

3

a describe the portfolio approach to investing;

In the previous section we compared an equally weighted portfolio to the selection of a single security. In this section we examine additional combinations of the same set of shares and observe the trade-offs between portfolio volatility of returns and expected return (for short, their risk–return trade-offs). If we select the portfolios with the best combination of risk and return (taking historical statistics as our expectations for the future), we produce the set of portfolios shown in Exhibit 5.

Exhibit 5 Optimal Portfolios for Sample of HKSE Listed Shares



Source: Thomson Reuters Datastream.

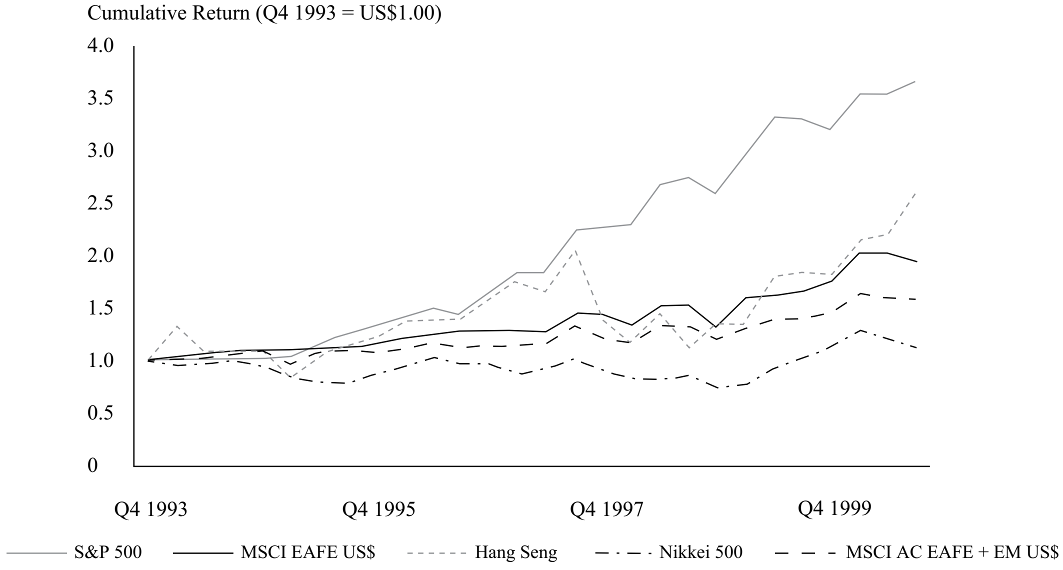
In addition to illustrating that the diversified portfolio approach reduces risk, Exhibit 5 also shows that the composition of the portfolio matters. For example, an equally weighted portfolio (20 percent of the portfolio in each security) of the five shares has an expected return of 15.1 percent and a standard deviation of 17.9 percent. Alternatively, a portfolio with 25 percent in Yue Yuen Industrial (Holdings), 3 percent in Cathay Pacific, 52 percent in Hutchison Whampoa, 20 percent in Li & Fung, and 0 percent in COSCO Pacific produces a portfolio with an expected return of 15.1 percent and a standard deviation of 15.6 percent. Compared to a simple equally weighted portfolio, this provides an improved trade-off between risk and return because a lower level of risk was achieved for the same level of return.

3.1 Historical Portfolio Example: Not Necessarily Downside Protection

A major reason that portfolios can effectively reduce risk is that combining securities whose returns do not move together provides diversification. Sometimes a subset of assets will go up in value at the same time that another will go down in value. The fact that these may offset each other creates the potential diversification benefit we attribute to portfolios. However, an important issue is that the co-movement or correlation pattern of the securities' returns in the portfolio can change in a manner unfavorable to the investor. We use historical return data from a set of global indexes to show the impact of changing co-movement patterns.

When we examine the returns of a set of global equity indexes over the last 15 years, we observe a reduction in the diversification benefit due to a change in the pattern of co-movements of returns. Exhibits 6 and 7 show the cumulative returns for a set of five global indexes[7] for two different time periods. Comparing the first time period, from Q4 1993 through Q3 2000 (as shown in Exhibit 6), with the last time period, from Q1 2006 through Q1 2009 (as shown in Exhibit 7), we show that the degree to which these global equity indexes moved together increased over time.

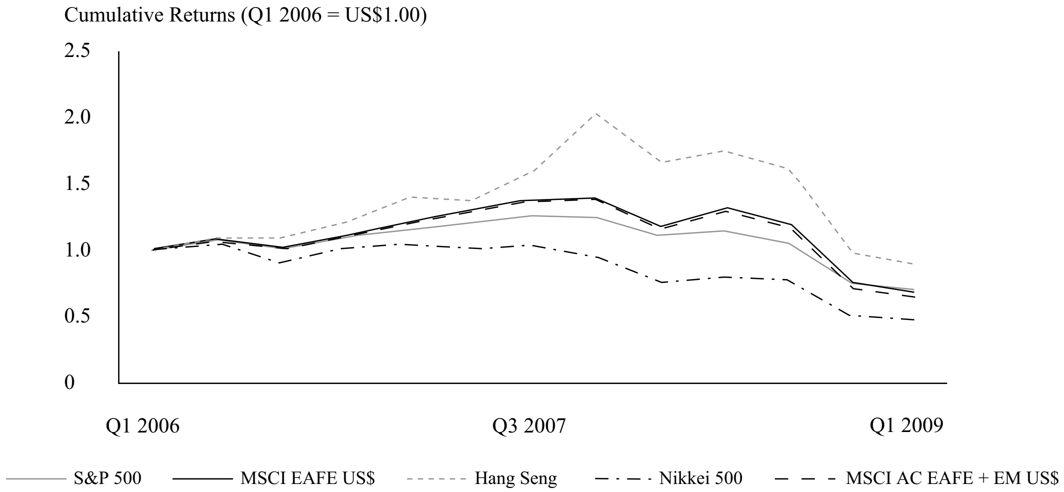
Exhibit 6 Returns to Global Equity Indexes Q4 1993–Q3 2000



Source: Thomson Reuters Datastream.

7 The S&P 500, Hang Seng, and Nikkei 500 are broad-based composite equity indexes designed to measure the performance of equities in the United States, Hong Kong SAR, and Japan. MSCI stands for Morgan Stanley Capital International. EAFE refers to developed markets in Europe, Australasia, and the Far East. AC indicates all countries, and EM is emerging markets. All index returns are in US dollars.

Exhibit 7 Returns to Global Equity Indexes Q1 2006–Q1 2009



Source: Thomson Reuters Datastream.

The latter part of the second time period, from Q4 2007 to Q1 2009, was a period of dramatic declines in global share prices. Exhibit 8 shows the mean annual returns and standard deviation of returns for this time period.

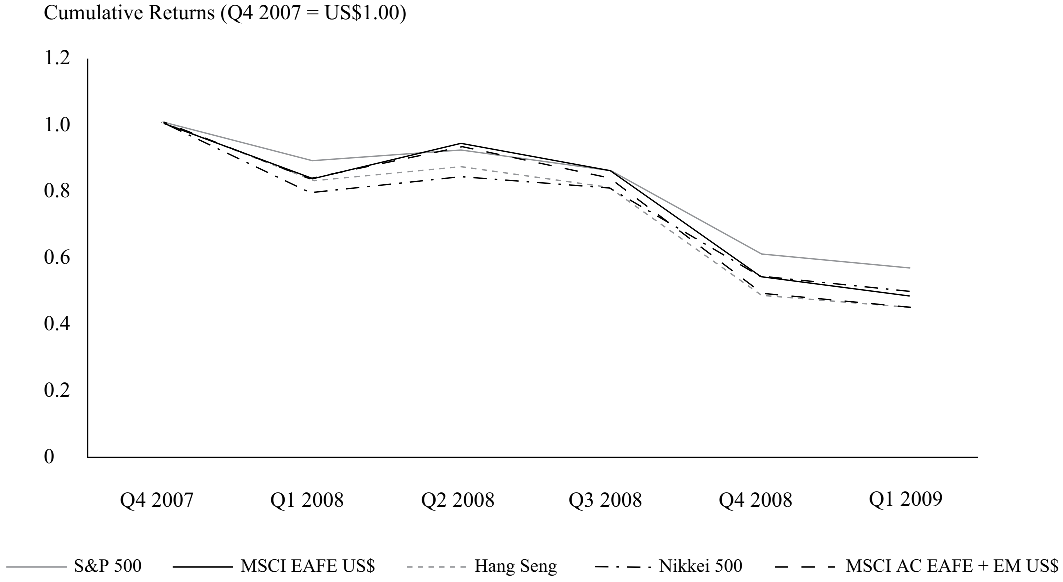
Exhibit 8 Returns to Global Equity Indexes

	Q4 1993–Q3 2000		Q1 2006–Q1 2009		Q4 2007–Q1 2009	
Global Index	**Mean**	**Stand. Dev.**	**Mean**	**Stand. Dev.**	**Mean**	**Stand. Dev.**
S&P 500	20.5%	13.9%	−6.3%	21.1%	−40.6%	23.6%
MSCI EAFE US$	10.9	14.2	−3.5	29.4	−48.0	35.9
Hang Seng	20.4	35.0	5.1	34.2	−53.8	34.0
Nikkei 500	3.3	18.0	−13.8	27.6	−48.0	30.0
MSCI AC EAFE + EM US$	7.6	13.2	−4.9	30.9	−52.0	37.5
Randomly selected index	12.6%	18.9%	−4.7%	28.6%	−48.5%	32.2%
Equally weighted portfolio	12.6%	14.2%	−4.7%	27.4%	−48.5%	32.0%
Diversification ratio		75.1%		95.8%		99.4%

Source: Thomson Reuters Datastream.

During the period Q4 2007 through Q1 2009, the average return for the equally weighted portfolio, including dividends, was −48.5 percent. Other than reducing the risk of earning the return of the worst performing market, the diversification benefits were small. Exhibit 9 shows the cumulative quarterly returns of each of the five indexes over this time period. All of the indexes declined in unison. The lesson is that although portfolio diversification generally does reduce risk, it does not necessarily provide the same level of risk reduction during times of severe market turmoil as it does when the economy and markets are operating 'normally'. In fact, if the economy or markets fail totally (which has happened numerous times around the world), then diversification is a false promise. In the face of a worldwide contagion, diversification was ineffective, as illustrated at the end of 2008.

Exhibit 9 Return to Global Equity Indexes Q4 2007–Q1 2009



Source: Thomson Reuters Datastream.

Portfolios are *most likely* to provide:

A risk reduction.

B risk elimination.

C downside protection.

Solution:

A is correct. Combining assets into a portfolio should reduce the portfolio's volatility. However, the portfolio approach does not necessarily provide downside protection or eliminate all risk.

3.2 Portfolios: Modern Portfolio Theory

The concept of diversification has been around for a long time and has a great deal of intuitive appeal. However, the actual theory underlying this basic concept and its application to investments only emerged in 1952 with the publication of Harry Markowitz's classic article on portfolio selection.[8] The article provided the foundation for what is now known as **modern portfolio theory** (MPT). The main conclusion of MPT is that investors should not only hold portfolios but should also focus on how individual securities in the portfolios are related to one another. In addition to the

8 Markowitz (1952).

diversification benefits of portfolios to investors, the work of William Sharpe (1964), John Lintner (1965), and Jack Treynor (1961) demonstrated the role that portfolios play in determining the appropriate individual asset risk premium (i.e., the return in excess of the risk-free return expected by investors as compensation for the asset's risk). According to capital market theory, the priced risk of an individual security is affected by holding it in a well-diversified portfolio. The early research provided the insight that an asset's risk should be measured in relation to the remaining systematic or non-diversifiable risk, which should be the only risk that affects the asset's price. This view of risk is the basis of the capital asset pricing model, or CAPM, which is discussed in greater detail in other readings. Although MPT has limitations, the concepts and intuitions illustrated in the theory continue to be the foundation of knowledge for portfolio managers.

4 STEPS IN THE PORTFOLIO MANAGEMENT PROCESS

b describe the steps in the portfolio management process;

In the previous section we discussed a portfolio approach to investing. When establishing and managing a client's investment portfolio, certain critical steps are followed in the process. We describe these steps in this section.

- The Planning Step
  - Understanding the client's needs
  - Preparation of an investment policy statement (IPS)
- The Execution Step
  - Asset allocation
  - Security analysis
  - Portfolio construction
- The Feedback Step
  - Portfolio monitoring and rebalancing
  - Performance measurement and reporting

4.1 Step One: The Planning Step

The first step in the investment process is to understand the client's needs (objectives and constraints) and develop an **investment policy statement** (IPS). A portfolio manager is unlikely to achieve appropriate results for a client without a prior understanding of the client's needs. The IPS is a written planning document that describes the client's investment objectives and the constraints that apply to the client's portfolio. The IPS may state a benchmark—such as a particular rate of return or the performance of a particular market index—that can be used in the feedback stage to assess the performance of the investments and whether objectives have been met. The IPS should be reviewed and updated regularly (for example, either every three years or when a major change in a client's objectives, constraints, or circumstances occurs).

4.2 Step Two: The Execution Step

The next step is for the portfolio manager to construct a suitable portfolio based on the IPS of the client. The portfolio execution step consists of first deciding on a target asset allocation, which determines the weighting of asset classes to be included in the portfolio. This step is followed by the analysis, selection, and purchase of individual investment securities.

4.2.1 *Asset Allocation*

The next step in the process is to assess the risk and return characteristics of the available investments. The analyst forms economic and capital market expectations that can be used to form a proposed allocation of asset classes suitable for the client. Decisions that need to be made in the **asset allocation** of the portfolio include the distribution between equities, fixed-income securities, and cash; sub-asset classes, such as corporate and government bonds; and geographical weightings within asset classes. Alternative assets—such as real estate, commodities, hedge funds, and private equity—may also be included.

Economists and market strategists may set the top down view on economic conditions and broad market trends. The returns on various asset classes are likely to be affected by economic conditions; for example, equities may do well when economic growth has been unexpectedly strong whereas bonds may do poorly if inflation increases. The economists and strategists will attempt to forecast these conditions.

Top down—A **top-down analysis** begins with consideration of macroeconomic conditions. Based on the current and forecasted economic environment, analysts evaluate markets and industries with the purpose of investing in those that are expected to perform well. Finally, specific companies within these industries are considered for investment.

Bottom up—Rather than emphasizing economic cycles or industry analysis, a **bottom-up analysis** focuses on company-specific circumstances, such as management quality and business prospects. It is less concerned with broad economic trends than is the case for top-down analysis, but instead focuses on company specifics.

4.2.2 *Security Analysis*

The top-down view can be combined with the bottom-up insights of security analysts who are responsible for identifying attractive investments in particular market sectors. They will use their detailed knowledge of the companies and industries they cover to assess the expected level and risk of the cash flows that each security will produce. This knowledge allows the analysts to assign a valuation to the security and identify preferred investments.

4.2.3 *Portfolio Construction*

The portfolio manager will then construct the portfolio, taking account of the target asset allocation, security analysis, and the client's requirements as set out in the IPS. A key objective will be to achieve the benefits of diversification (i.e., to avoid putting all the eggs in one basket). Decisions need to be taken on asset class weightings, sector weightings within an asset class, and the selection and weighting of individual securities or assets. The relative importance of these decisions on portfolio performance depends at least in part on the investment strategy selected; for example, consider an investor that actively adjusts asset sector weights in relation to forecasts of sector performance and one who does not. Although all decisions have an effect on portfolio performance, the asset allocation decision is commonly viewed as having the greatest impact.

Exhibit 10 shows the broad portfolio weights of the endowment funds of Yale University and the University of Virginia as of June 2017. As you can see, the portfolios have a heavy emphasis on such alternative assets as hedge funds, private equity, and real estate—Yale University particularly so.

Exhibit 10 Endowment Portfolio Weights, June 2017

Asset Class	Yale University Endowment	University of Virginia Endowment
Public equity	19.1%	26.7%
Fixed income	4.6	9.1
Private equity	14.2	15.7
Real assets (e.g., real estate)	18.7	12.1
Absolute return (e.g., hedge funds)	25.1	19.6
Cash	1.2	2.3
Other	17.2	14.5
Portfolio value	US$27.2bn	US$8.6bn

Sources: "2017 Yale Endowment Annual Report" (p. 2): www.yale.edu/investments/Yale_Endowment_17.pdf; "University of Virginia Investment Management Company Annual Report 2017" (p. 26): http://uvm-web.eservices.virginia.edu/public/reports/FinancialStatements_2017.pdf.

Risk management is an important part of the portfolio construction process. The client's risk tolerance will be set out in the IPS, and the portfolio manager must make sure the portfolio is consistent with it. As noted above, the manager will take a diversified portfolio perspective: What is important is not the risk of any single investment, but rather how all the investments perform as a portfolio.

The endowments shown above are relatively risk tolerant investors. Contrast the asset allocation of the endowment funds with the portfolio mix of the insurance companies shown in Exhibit 11. You will notice that the majority of the insurance assets are invested in fixed-income investments, typically of high quality. Note that the Yale University portfolio has less than 5 percent invested in fixed income, with the remainder invested in such growth assets as equity, real estate, and hedge funds. This allocation is in sharp contrast to the Massachusetts Mutual Life Insurance Company (MassMutual) portfolio, which is 80 percent invested in bonds, mortgages, loans, and cash—reflecting the differing risk tolerance and constraints (life insurers face regulatory constraints on their investments).

Exhibit 11 MassMutual Portfolio, December 2017[9]

Asset Classes	Portfolio %
Bonds	56%
Preferred and common shares	9
Mortgages	14
Real estate	1
Policy loans	8
Partnerships	5
Other assets	5
Cash	2

Source: "MassMutual Financial Group 2017 Annual Report" (p. 8): www.massmutual.com/mmfg/docs/annual_report/index.html.

The portfolio construction phase also involves trading. Once the portfolio manager has decided which securities to buy and in what amounts, the securities must be purchased. In many investment firms, the portfolio manager will pass the trades to a buy-side trader—a colleague who specializes in securities trading—who will contact a stockbroker or dealer to have the trades executed.

4.3 Step Three: The Feedback Step

Finally, the feedback step assists the portfolio manager in rebalancing the portfolio due to a change in, for example, market conditions or the circumstances of the client.

4.3.1 *Portfolio Monitoring and Rebalancing*

Once the portfolio has been constructed, it needs to be monitored and reviewed and the composition revised as the security analysis changes because of changes in security prices and changes in fundamental factors. When security and asset weightings have drifted from the intended levels as a result of market movements, some rebalancing may be required. The portfolio may also need to be revised if it becomes apparent that the client's needs or circumstances have changed.

4.3.2 *Performance Evaluation and Reporting*

Finally, the performance of the portfolio must be evaluated, which will include assessing whether the client's objectives have been met. For example, the investor will wish to know whether the return requirement has been achieved and how the portfolio has performed relative to any benchmark that has been set. Analysis of performance may suggest that the client's objectives need to be reviewed and perhaps changes made to the IPS. As we will discuss in the next section, there are numerous investment products that clients can use to meet their investment needs. Many of these products are diversified portfolios that an investor can purchase.

9 Asset class definitions: Bonds—Debt instruments of corporations and governments as well as various types of mortgage- and asset-backed securities; Preferred and Common Shares—Investments in preferred and common equities; Mortgages—Mortgage loans secured by various types of commercial property as well as residential mortgage whole loan pools; Real Estate—Investments in real estate; Policy Loans—Loans by policyholders that are secured by insurance and annuity contracts; Partnerships—Investments in partnerships and limited liability companies; Cash—Cash, short-term investments, receivables for securities, and derivatives. Cash equivalents have short maturities (less than one year) or are highly liquid and able to be readily sold.

TYPES OF INVESTORS

5

c describe types of investors and distinctive characteristics and needs of each;

d describe defined contribution and defined benefit pension plans;

The portfolio management process described in the previous section may apply to different types of investment clients. Such clients are broadly divided among individual (or retail) and institutional investors. Each of these segments has distinctive characteristics and needs, as discussed in the following sub-sections.

5.1 Individual Investors

Individual investors have a variety of motives for investing and constructing portfolios. Short-term goals can include providing for children's education, saving for a major purchase (such as a vehicle or a house), or starting a business. The retirement goal—investing to provide for an income in retirement—is a major part of the investment planning of most individuals. Many employees of public and private companies invest for retirement through **defined contribution pension plans** (DC plans). DC plans are retirement plans in the employee's name usually funded by both the employee and the employer. Examples include 401(k) plans in the United States, group personal pension schemes in the United Kingdom, and superannuation plans in Australia. With DC plans, individuals will invest part of their wages while working, expecting to draw on the accumulated funds to provide income during retirement or to transfer some of their wealth to their heirs. The key to a DC plan is that the employee accepts the investment and inflation risk and is responsible for ensuring that there are enough assets in the plan to meet their needs upon retirement.

Some individuals will be investing for growth and will therefore seek assets that have the potential for capital gains. Others, such as retirees, may need to draw an income from their assets and may therefore choose to invest more in fixed-income and dividend-paying shares. The investment needs of individuals will depend in part on their broader financial circumstances, such as their employment prospects and whether or not they own their own residence. They may also need to consider such issues as building up a cash reserve and the purchase of appropriate insurance policies before undertaking longer-term investments.

Asset managers serving individual investors typically distribute their products directly to investors or through intermediaries such as financial advisers and/or retirement plan providers. The distribution network for individual investors varies globally. In the United States, financial advisers are independent or employed by national or regional broker–dealers, banks, and trust companies. Additionally, many asset managers distribute investment strategies to investors through major online brokerage and custodial firms.

In Europe, retail investment product distribution is fragmented and, in turn, varies by country/region. In continental Europe, for example, distribution is primarily driven through financial advisers affiliated with retail and private banks. In the United Kingdom, products are sold through independent advisers as well as through advisers representing a bank or insurance group. Retail distribution in Switzerland and in the Nordic countries is driven mainly through large regional and private banks. In contrast to the United States and Europe, in many Asian markets retail distribution is dominated by large regional retail banks and global banks with private banking divisions.

Globally, many wealth management firms and asset managers target high-net-worth investors. These clients often require more customized investment solutions alongside tax and estate planning services.

5.2 Institutional Investors

Institutional investors primarily include defined benefit pension plans, endowments and foundations, banks, insurance companies, investment companies, and sovereign wealth funds. Each of these has unique goals, asset allocation preferences, and investment strategy needs.

5.2.1 *Defined Benefit Pension Plans*

Pension plans are typically categorized as either defined contribution (DC) or defined benefit (DB). We previously described DC plans, which relate to individual investors. **Defined benefit pension plans** (DB plans) are company-sponsored plans that offer employees a predefined benefit on retirement. The future benefit is defined because the DB plan requires the plan sponsor to specify the obligation stated in terms of the retirement income benefits owed to participants. Generally, employers are responsible for the contributions made to a DB plan and bear the risk associated with adequately funding the benefits offered to employees. Plans are committed to paying pensions to members, and the assets of these plans are there to fund those payments. Plan managers need to ensure that sufficient assets will be available to pay pension benefits as they come due. The plan may have an indefinitely long time horizon if new plan members are being admitted or a finite time horizon if the plan has been closed to new members. In some cases, the plan managers attempt to match the fund's assets to its liabilities by, for example, investing in bonds that will produce cash flows corresponding to expected future pension payments. There may be many different investment philosophies for pension plans, depending on funded status and other variables.

An ongoing trend is that plan sponsors increasingly favor DC plans over DB plans because DC plans typically have lower costs/risk to the company. As a result, DB plans have been losing market share of pension assets to DC plans. Nevertheless, DB plans, both public and private, remain sizable sources of investment funds for asset managers. As Exhibit 12 shows, global pension assets totaled more than US$41 trillion by the end of 2017. The United States, United Kingdom, and Japan represent the three largest pension markets in the world, comprising more than 76% of global pension assets.

Exhibit 12 Global Pension Assets (as of year-end 2017)

Country/Region	Total Assets (US$ billions)
United States	25,411
United Kingdom	3,111
Japan	3,054
Australia	1,924
Canada	1,769
Netherlands	1,598
Switzerland	906
South Korea	725
Germany	472
Brazil	269
South Africa	258
Finland	233
Malaysia	227
Chile	205
Mexico	177

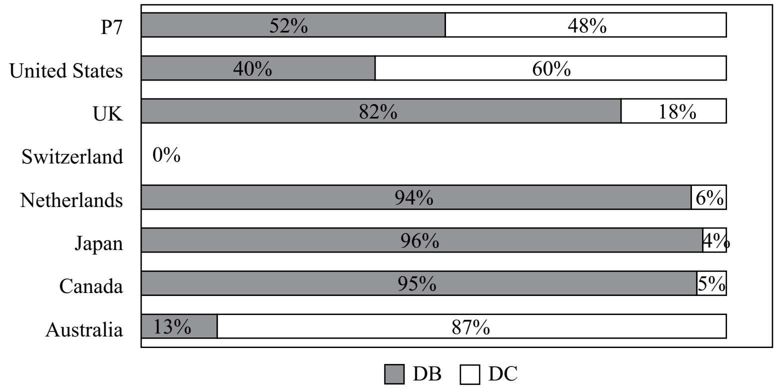
Exhibit 12 (Continued)

Country/Region	Total Assets (US$ billions)
Italy	184
France	167
Chinese mainland	177
Hong Kong SAR	164
Ireland	157
India	120
Spain	44
Total	**41,355**

Note: Column does not sum precisely because of rounding.
Source: Willis Towers Watson.

By geography, the United States and Australia have a higher proportion of pension assets in DC plans, whereas Canada, Japan, the Netherlands, and the United Kingdom remain weighted toward DB plans (see Exhibit 13).

Exhibit 13 Pension Plan Type by Geography



Notes: "P7" represents the combination of the seven countries listed. No data were available for Switzerland for this study.
Sources: Willis Towers Watson and secondary sources.

5.2.2 *Endowments and Foundations*

Endowments are funds of non-profit institutions that help the institutions provide designated services. In contrast, foundations are grant-making entities. Endowments and foundations collectively represent an estimated US$1.6 trillion in assets in the United States, which is the primary market for endowments and foundations.

Endowments and foundations typically allocate a sizable portion of their assets in alternative investments (Exhibit 14). This large allocation to alternative investments primarily reflects the typically long time horizon of endowments and foundations, as well the popularity of endowment-specific asset allocation models developed by Yale University's endowment managers David Swensen and Dean Takahashi.

Exhibit 14 Asset Allocations for US College and University Endowments and Affiliated Foundations (as of 30 June 2017, dollar weighted)

Asset Class	Percentage Allocation
Domestic equity	15
Fixed income	7
Foreign equity	20
Alternatives	54
Cash	4

Source: National Association of College and University Budget Officers and Commonfund Institute.

A typical investment objective of an endowment or a foundation is to maintain the real (inflation-adjusted) capital value of the fund while generating income to fund the objectives of the institution. Most foundations and endowments are established with the intent of having perpetual lives. Exhibit 15 describes the Yale University endowment's approach to balancing short-term spending needs with ensuring that future generations also benefit from the endowment, and it also shows the Wellcome Trust's approach. The investment approach undertaken considers the objectives and constraints of the institution (for example, no tobacco investments for a medical endowment).

Exhibit 15 Spending Rules

The following examples of spending rules are excerpts from the Yale University endowment (in the United States) and from the Wellcome Trust (in the United Kingdom).

Yale University Endowment

> The spending rule is at the heart of fiscal discipline for an endowed institution. Spending policies define an institution's compromise between the conflicting goals of providing substantial support for current operations and preserving purchasing power of Endowment assets. The spending rule must be clearly defined and consistently applied for the concept of budget balance to have meaning.
>
> The Endowment spending policy, which allocates Endowment earnings to operations, balances the competing objectives of providing a stable flow of income to the operating budget and protecting the real value of the Endowment over time. The spending policy manages the trade-of between these two objectives by combining a long-term spending rate target with a smoothing rule, which adjusts spending in any given year gradually in response to changes in Endowment market value.
>
> The target spending rate approved by the Yale Corporation currently stands at 5.25%. According to the smoothing rule, Endowment spending in a given year sums to 80% of the previous year's spending and 20% of the targeted long-term spending rate applied to the fiscal year-end market value two years prior.

Exhibit 15 (Continued)

Source: 2017 Yale Endowment Annual Report (p.18) [http://investments.yale.edu/endowment-update/]

Wellcome Trust

Our overall investment objective is to generate 4.5% percent real return over the long term.

This is to provide for real increases in annual expenditure while reserving the Trust's capital base to balance the needs of current and future beneficiaries.

We use this absolute return strategy because it aligns asset allocation with funding requirements and provides a competitive framework in which to judge individual investments.

Source: Wellcome Trust website (https://wellcome.ac.uk/about-us/investments)

5.2.3 *Banks*

Banks are financial intermediaries that accept deposits and lend money. Banks often have excess reserves that are invested in relatively conservative and very short-duration fixed-income investments, with a goal of earning an excess return above interest obligations due to depositors. Liquidity is a paramount concern for banks that stand ready to meet depositor requests for withdrawals. Many large banks have asset management divisions that offer retail and institutional products to their clients.

5.2.4 *Insurance Companies*

Insurance companies receive premiums for the policies they write, and they need to invest these premiums in a manner that will allow them to pay claims.

Insurance companies can be segmented into two broad types: life insurers and property and casualty (P&C) insurers. Insurance premiums from policyholders comprise an insurance company's general account. To pay claims to policyholders, regulatory guidelines maintain that an insurance company's general account is typically invested conservatively in a diverse allocation of fixed-income securities. General account portfolio allocations differ among life, P&C, and other specialty insurers (e.g., reinsurance) because of both the varying duration of liabilities and the unique liquidity considerations across insurance type.[10] In contrast to the general account, an insurer's surplus account is the difference between its assets and liabilities. An insurer's surplus account typically targets a higher return than the general account and thus often invests in less-conservative asset classes, such as public and private equities, real estate, infrastructure, and hedge funds.

Many insurance companies have in-house portfolio management teams responsible for managing general account assets. Some insurance companies offer portfolio management services and products in addition to their insurance offerings. An increasing trend among insurers (particularly in the United States) is outsourcing some of the

10 For example, life insurers tend to invest in longer-term assets (e.g., 30-year government and corporate bonds) relative to P&C insurers because of the longer-term nature of their liabilities.

portfolio management responsibilities—primarily sophisticated alternative asset classes—to unaffiliated asset managers. Several insurers manage investments for third-party clients, often through separately branded subsidiaries.

5.2.5 *Sovereign Wealth Funds*

Sovereign wealth funds (SWFs) are state-owned investment funds or entities that invest in financial or real assets. SWFs do not typically manage specific liability obligations, such as pensions, and have varying investment horizons and objectives based on funding the government's goals (for example, budget stabilization or future development projects). SWF assets more than doubled from 2007 to March 2018, totaling more than US$7.6 trillion.[11] Exhibit 16 lists the 10 largest SWFs in the world. The largest SWFs tend to be concentrated in Asia and in natural resource-rich places.

Exhibit 16 Largest Sovereign Wealth Funds (as of August 2018, in US$ billions)

Place	Sovereign Wealth Fund (Inception Year)	Assets
Norway	Government Pension Fund—Global (1990)	1,058
Chinese Mainland	China Investment Corporation (2007)	941
UAE – Abu Dhabi	Abu Dhabi Investment Authority (1976)	683
Kuwait	Kuwait Investment Authority (1953)	592
Hong Kong SAR	Hong Kong Monetary Authority Investment Portfolio (1993)	523
Saudi Arabia	SAMA Foreign Holdings (1952)	516
Chinese Mainland	SAFE Investment Company (1997)	441
Singapore	Government of Singapore Investment Authority (1981)	390
Singapore	Temasek Holdings (1974)	375
Saudi Arabia	Public Investment Fund (2008)	360
Total SWF Assets under Management		**8,109**

Source: SWF Institute (www.swfinstitute.org).

Investment needs vary across client groups. With some groups of clients, generalizations are possible. In other groups, needs vary by client. Exhibit 17 summarizes needs within each group.

Exhibit 17 Summary of Investment Needs by Client Type

Client	Time Horizon	Risk Tolerance	Income Needs	Liquidity Needs
Individual investors	Varies by individual	Varies by individual	Varies by individual	Varies by individual
Defined benefit pension plans	Typically long term	Typically quite high	High for mature funds; low for growing funds	Varies by maturity of the plan

11 SWFI, "Sovereign Wealth Fund Rankings" (https://www.swfinstitute.org/sovereign-wealth-fund-rankings/; retrieved October 2018).

Exhibit 17 (Continued)

Client	Time Horizon	Risk Tolerance	Income Needs	Liquidity Needs
Endowments and foundations	Very long term	Typically high	To meet spending commitments	Typically quite low
Banks	Short term	Quite low	To pay interest on deposits and operational expenses	High to meet repayment of deposits
Insurance companies	Short term for property and casualty; long term for life insurance companies	Typically quite low	Typically low	High to meet claims
Investment companies	Varies by fund	Varies by fund	Varies by fund	High to meet redemptions
Sovereign wealth funds	Varies by fund	Varies by fund	Varies by fund	Varies by fund

THE ASSET MANAGEMENT INDUSTRY

6

e describe aspects of the asset management industry;

The portfolio management process and investor types are broad components of the asset management industry, which is an integral component of the global financial services sector. At the end of 2017, the industry managed more than US$79 trillion of assets owned by a broad range of institutional and individual investors (Exhibit 18).[12] Although nearly 80% of the world's professionally managed assets are in North America and Europe, the fastest-growing markets are in Asia and Latin America.

Exhibit 18 Global Assets under Management (AUM) by Region (year-end 2017)

	Market Size (US$ trillions)	Market Share (%)
North America	37.4	47%
Europe	22.2	28
Japan and Australia	6.2	8
Chinese mainland	4.2	5
Asia (excluding Japan, Australia, and Chinese mainland)	3.5	4
Latin America	1.8	2
Middle East and Africa	1.4	2
Total Global AUM	79.2	100%

Notes: Total Global AUM in this exhibit represents assets professionally managed in exchange for a fee. The total of US$79.2 trillion includes certain offshore assets that are not represented in the specific regional categories above.

(continued)

12 http://image-src.bcg.com/Images/BCG-Seizing-the-Analytics-Advantage-June-2018-R-3_tcm9-194512.pdf (accessed on 6 September 2018).

Exhibit 18 (Continued)

Source: Boston Consulting Group.

The asset management industry is highly competitive. The universe of firms in the industry is broad, ranging from "pure-play" independent asset managers to diversified commercial banks, insurance companies, and brokerages that offer asset management services in addition to their core business activities. Given the increasingly global nature of the industry, many asset managers have investment research and distribution offices around the world. An asset manager is commonly referred to as a **buy-side firm** given that it uses (buys) the services of sell-side firms. A **sell-side firm** is a broker/dealer that sells securities and provides independent investment research and recommendations to their clients (i.e., buy-side firms).

Asset managers offer a broad range of strategies. Specialist asset managers may focus on a specific asset class (e.g., emerging market equities) or style (e.g., quantitative investing), while "full service" managers typically offer a wide variety of asset classes and styles. Another type of asset manager firm is a "multi-boutique," in which a holding company owns several asset management firms that typically have specialized investment strategies. The multi-boutique structure allows individual asset management firms to retain their own unique investment cultures—and often equity ownership stakes—while also benefiting from the centralized, shared services of the holding company (e.g., technology, sales and marketing, operations, and legal services).

6.1 Active versus Passive Management

Asset managers may offer either active or passive management. As of year-end 2017, active management considerably exceeded passive management in terms of global assets under management and industry revenue (Exhibit 19), although passive management has demonstrated significant growth.

Exhibit 19 Global Asset Management Industry Assets and Revenue (as of year-end 2017)

Category	Assets (US$ trillions)	Revenue (US$ billions)	Market Share by Assets (%)	Market Share by Revenue (%)
Actively Managed	**64**	**258**	**80%**	**94%**
Alternatives	12	117	15	43
Active Specialties	15	55	19	20
Multi Asset Class	11	27	14	10
Core	26	59	33	21
Passively Managed	**16**	**17**	**20%**	**6%**
Total	**80**	**275**	**100%**	**100%**

Note: Some columns may not sum precisely because of rounding.
Source: Boston Consulting Group.

Through fundamental research, quantitative research, or a combination of both, active asset managers generally attempt to outperform either predetermined performance benchmarks, such as the S&P 500, or, for multi-asset class portfolios, a combination of benchmarks. In contrast to active managers, passive managers attempt to replicate the returns of a market index. Despite the rise of passive management in asset share, its share of industry revenue remains small given the low management fees relative to active management. As Exhibit 19 illustrates, passive management represents a fifth of global assets under management but only 6% of industry revenue.

Asset managers are increasingly offering other strategies beyond traditional market-cap-weighted exposures. Some of these other strategies, commonly known as **smart beta**, are based on such factors as size, value, momentum, or dividend characteristics. Smart beta involves the use of simple, transparent, rules-based strategies as a basis for investment decisions. Typically, smart beta strategies feature somewhat higher management fees and higher portfolio turnover relative to passive market-cap weighted strategies.

6.2 Traditional versus Alternative Asset Managers

Asset managers are typically categorized as either "traditional" or "alternative." Traditional managers generally focus on long-only equity, fixed-income, and multi-asset investment strategies, generating most of their revenues from asset-based management fees. Alternative asset managers, however, focus on hedge fund, private equity, and venture capital strategies, among others, while generating revenue from both management and performance fees (or "carried interest"). As Exhibit 19 demonstrates, alternative asset managers have a relatively low proportion of total global assets under management but generate a disproportionately high total of industry revenue.

Increasingly, the line between traditional and alternative managers has blurred. Many traditional managers have introduced higher-margin alternative products to clients. Concurrently, alternative managers seeking to reduce the revenue volatility associated with performance fees have increasingly offered retail versions of their institutional alternative strategies (typically referred to as "liquid alternatives") as well as long-only investment strategies. These liquid alternatives are often offered through highly regulated pooled investment products (e.g., mutual funds) and typically feature less leverage, no performance fees, and more liquid holdings than typical alternative products.

6.3 Ownership Structure

The ownership structure of an asset manager can play an important role in retaining and incentivizing key personnel. Portfolio managers who have personal capital invested in their firms or investment strategies are often viewed favorably by potential investors because of perceived alignment of management and client interests.

The majority of asset management firms are privately owned, typically by individuals who either established their firms or play key roles in their firms' management. Privately owned firms are typically structured as limited liability companies or limited partnerships.

While less common than privately owned managers, publicly traded asset managers have substantial assets under management. A prevalent ownership form in the industry is represented by asset management divisions of large, diversified financial services companies that offer asset management alongside insurance and banking services.

6.4 Asset Management Industry Trends

The asset management industry is evolving and continues to be shaped by socio-economic trends, shifting investor demands, advances in technology, and the expansion of global capital markets. Three key trends that we discuss in this section include the growth of passive investing, "big data" in the investment process, and the emergence of robo-advisers in the wealth management industry.

6.4.1 *Growth of Passive Investing*

As we saw in Exhibit 19, passively managed assets comprised nearly a fifth of global assets under management at the end of 2017. Management of passive assets is concentrated among a reasonably small group of asset managers and tends to be concentrated in equity strategies. As shown in Exhibit 20, the top three managers account for 70% of industry's assets. One key catalyst supporting the growth of passive investing is low cost for investors—management fees for index (or other passive) funds are often a fraction of those for active strategies. Another catalyst is the challenge that many active asset managers face in generating ex ante alpha, especially in somewhat more-efficiently priced markets, such as large-cap US equities.

Exhibit 20 Top Five ETP Managers Globally (as of 30 July 2017)

ETP Provider	Assets (US$ billions)	Market Share (%)
iShares	1,583	37
Vanguard	803	19
State Street Global Advisors	596	14
PowerShares	132	3
Nomura	100	2

Source: ETFGI.

6.4.2 *Use of "Big Data" in the Investment Process*

The prevalence of new data is extraordinary: In 2013, IBM estimated that 90% of the world's entire universe of data was created in the previous two years. The digitization of data and an exponential increase in computing power and data storage capacity have expanded additional information sources for asset managers. Massive amounts of data containing information of potential value to investors are created and captured daily. These data include both structured data—such as order book data and security returns—and data lacking recognizable structure, which is generated by a vast number of activities on the internet and elsewhere (e.g., compiled search information). The term "big data" is used to refer to these massively large datasets and their analysis.

Asset managers are using advanced statistical and machine-learning techniques to help process and analyze these new sources of data. Such techniques are used in both fundamentally driven and quantitatively driven investment processes. For example, computers are used to "read" earnings and economic data releases much faster than humans can and react with short-term trading strategies.

Third-party research vendors are supplying a vast range of relevant new data for asset managers, such as data used for time-series and predictive models. Among the most popular new sources of data are social media data and imagery and sensor data.

- **Social media data**. Real-time media and content outlets, such as Twitter and Facebook, provide meaningful market and company-specific announcements for investors and asset managers. In addition, the aggregation and analysis of social media users can aid key market sentiment indicators (e.g., short-term directional market movements) and indicate potential specific user trends related to products and services.

- **Imagery and sensor data**. Satellite imagery and geolocation devices provide vast real-time data to investment professionals. As the cost of launching and maintaining satellites has decreased, more satellites have been launched to track sensors and imagery that are relevant to economic considerations (e.g., weather conditions, cargo ship traffic patterns) and company-specific considerations (e.g., retailer parking capacity/usage, tracking of retail customers).

The challenge for asset managers is to discover data with predictive potential and to do so faster than fellow market participants. Many market participants are participating in an "information arms race" that has required substantial investments in specialized human capital (e.g., programmers, data scientists), technology, and information technology infrastructure to effectively convert various forms of structured and unstructured data into alpha-generating portfolio and security-level decisions.

6.4.3 *Robo-Advisers: An Expanding Wealth Management Channel*

Robo-advisers represent technology solutions that use automation and investment algorithms to provide several wealth management services—notably, investment planning, asset allocation, tax loss harvesting, and investment strategy selection. Investment and advice services provided by robo-advisers typically reflect an investor's general investment goals and risk tolerance preferences (often obtained from an investor questionnaire). Robo-adviser platforms range from exclusively digital investment advice platforms to hybrid offerings that offer both digital investment advice and the services of a human financial adviser.

At the end of 2017, robo-advisers managed an estimated US$180 billion in assets,[13] and market participants expect that number to grow considerably over time. This expected rapid growth in robo-advisory assets is based on several industry trends:

- **Growing demand from "mass affluent" and younger investors**: Traditional investment advice has often underserved younger and "mass affluent" investors with lower relative levels of investable assets. Given the efficiencies of robo-advisers and the scalability of technology, customized but standardized investment advice now can be offered to a wide range and size of investors.
- **Lower fees**: The cost of digital investment advice provided by robo-advisers is often a fraction of traditional investment advice channels because of scalability. For example, in the United States, a typical financial adviser may charge a 1% annual advisory fee[14] based on a client's assets, while robo-adviser fees typically average 0.20% annually.[15] Additionally, robo-advisers often rely on lower fee underlying portfolio investment options, such as index funds or ETFs, when constructing portfolios for clients.
- **New entrants**: Reflecting low barriers to entry, large wealth management firms have introduced robo-adviser solutions to service certain customer segments and appeal to a new generation of investors. In addition to these large wealth managers, other less-traditional entrants, such as insurance companies and asset managers, are developing solutions to cross-sell into their existing clients.

---

13 S&P Global Market Intelligence.

14 http://www.riainabox.com/blog/2016-ria-industry-study-average-investment-advisory-fee-is-0-99-percent.

15 Deloitte, "Robo-Advisors Capitalizing on a Growing Opportunity" (https://www2.deloitte.com/content/dam/Deloitte/us/Documents/strategy/us-cons-robo-advisors.pdf).

Many market observers expect that non-financial firms (large technology leaders) will also become key players in the robo-adviser industry as they look to monetize their access to user data.

7 POOLED INTEREST - MUTUAL FUNDS

f describe mutual funds and compare them with other pooled investment products.

In the asset management industry, a challenge faced by all investors is to and find the right set of investment products to meet their needs. There is a diverse set of investment products available to investors, ranging from a simple brokerage account in which the individual creates her own portfolio by assembling individual securities, to large institutions that employ individual portfolio managers to meet clients' investment management needs. Among the major investment products offered by asset managers are mutual funds and other pooled investment products, such as separately managed accounts, exchange-traded funds, hedge funds, and private equity/venture capital funds.

7.1 Mutual Funds

Rather than assemble a portfolio on their own, individual investors and institutions can turn over the selection and management of their investment portfolio to a third party. One way of doing this is through a **mutual fund**. This type of fund is a comingled investment pool in which investors in the fund each have a pro-rata claim on the income and value of the fund. The value of a mutual fund is referred to as the "net asset value." It is computed daily based on the closing price of the securities in the portfolio.

Mutual funds represent a primary investment product of individual investors globally. According to the International Investment Funds Association, worldwide regulated open-end fund assets totaled US$50 trillion as of the first quarter of 2018. Exhibit 21 shows the growth of global open-end funds over the past five years by region. Mutual funds provide several advantages, including low investment minimums, diversified portfolios, daily liquidity, and standardized performance and tax reporting.

Exhibit 21 Worldwide Regulated Open-End Funds: Total Net Assets (as of year-end, in US$ trillions)

	2011	2012	2013	2014	2015	2016	Q1 2018
World	**27.9**	**31.9**	**36.3**	**38.0**	**38.2**	**40.4**	**50.0**
Americas	14.6	16.5	18.9	20.0	19.6	21.1	24.9
Europe	10.3	11.9	13.6	13.8	13.7	14.1	18.1
Asia and Pacific	2.9	3.3	3.7	4.1	4.7	5.0	6.8
Africa	0.1	0.1	0.1	0.1	0.1	0.1	0.2

Notes: Components may not add to the total because of rounding. Regulated open-end funds include mutual funds, exchange-traded funds (ETFs), and institutional funds.
Source: International Investment Funds Association (IIFA).

Mutual funds are one of the most important investment vehicles for individuals and institutions. The best way to understand how a mutual fund works is to consider a simple example. Suppose that an investment firm wishes to start a mutual fund with a target amount of US$10 million. It is able to reach this goal through investments from five individuals and two institutions. The investment of each is as follows:

Investor	Amount Invested (US$)	Percent of Total	Number of Shares
Individuals			
A	$1.0 million	10%	10,000
B	1.0	10	10,000
C	0.5	5	5,000
D	2.0	20	20,000
E	0.5	5	5,000
Institutions			
X	2.0	20	20,000
Y	3.0	30	30,000
Totals	**$10.0 million**	**100%**	**100,000**

Based on the US$10 million value (net asset value), the investment firm sets a total of 100,000 shares at an initial value of US$100 per share (US$10 million/100,000 = US$100). The investment firm will appoint a portfolio manager to be responsible for the investment of the US$10 million. Going forward, the total value of the fund or net asset value will depend on the value of the assets in the portfolio.

The fund can be set up as an open-end fund or a closed-end fund. If it is an **open-end fund**, it will accept new investment money and issue additional shares at a value equal to the net asset value of the fund at the time of investment. For example, assume that at a later date the net asset value of the fund increases to US$12 million and the new net asset value per share is US$120. A new investor, F, wishes to invest US$0.96 million in the fund. If the total value of the assets in the fund is now US$12 million or US$120 per share, in order to accommodate the new investment the fund would create 8,000 (US$0.96 million/US$120) new shares. After this investment, the net asset value of the fund would be US$12.96 million and there would be a total of 108,000 shares.

Funds can also be withdrawn at the net asset value per share. Suppose on the same day Investor E wishes to withdraw all her shares in the mutual fund. To accommodate this withdrawal, the fund will have to liquidate US$0.6 million in assets to retire 5,000 shares at a net asset value of US$120 per share (US$0.6 million/US$120). The combination of the inflow and outflow on the same day would be as follows:

Type	Investment (US$)	Shares
Inflow (Investor F buys)	$960,000	8,000
Outflow (Investor E sells)	–$600,000	–5,000
Net	$360,000	3,000

The net of the inflows and outflows on that day would be US$360,000 of new funds to be invested and 3,000 new shares created. However, the number of shares held and the value of the shares of all remaining investors, except Investor E, would remain the same.

An alternative to setting the fund up as an open-end fund would be to create a **closed-end fund** in which no new investment money is accepted into the fund. New investors invest by buying existing shares, and investors in the fund liquidate by selling their shares to other investors. Hence, the number of outstanding shares does not change. One consequence of this fixed share base is that, unlike open-end funds in which new shares are created and sold at the current net asset value per share, closed-end funds can sell for a premium or discount to net asset value depending on the demand for the shares.

There are advantages and disadvantages to each type of fund. The open-end fund structure makes it easy to grow in size but creates pressure on the portfolio manager to manage the cash inflows and outflows. One consequence of this structure is the need to liquidate assets that the portfolio manager might not want to sell at the time to meet redemptions. Conversely, the inflows require finding new assets in which to invest. As such, open-end funds tend not to be fully invested but rather keep some cash for redemptions not covered by new investments. Closed-end funds do not have these problems, but they do have a limited ability to grow. Of the total net asset value of all US mutual funds at the end of 2017 (US$19 trillion), only approximately 1 percent were in the form of closed-end funds.

In addition to open-end or closed-end funds, mutual funds can be classified as load or no-load funds. The primary difference between the two is whether the investor pays a sales charge (a "load") to purchase, hold, or redeem shares in the fund. In the case of the **no-load fund**, there is no fee for investing in the fund or for redemption but there is an annual fee based on a percentage of the fund's net asset value. **Load funds** are funds in which, in addition to the annual fee, a percentage fee is charged to invest in the fund and/or for redemptions from the fund. In addition, load funds are usually sold through retail brokers who receive part of the upfront fee. Overall, the number and importance of load funds has declined over time.

8 POOLED INTEREST - TYPE OF MUTUAL FUNDS

f describe mutual funds and compare them with other pooled investment products.

The following section introduces the major types of mutual funds differentiated by the asset type that they invest in: money market funds, bond mutual funds, stock mutual funds, and hybrid or balanced funds.

8.1 Money Market Funds

Money market funds are mutual funds that invest in short-term money market instruments such as treasury bills, certificates of deposit, and commercial paper. They aim to provide security of principal, high levels of liquidity, and returns in line with money market rates. Many funds operate on a constant net asset value (CNAV) basis where the share price is maintained at $1 (or local currency equivalent). Others operate on a variable net asset value (VNAV) basis where the unit price can vary. In the United States, there are two basic types of money market funds: taxable and tax-free. Taxable money market funds invest in high-quality, short-term corporate debt and federal government debt. Tax-free money market funds invest in short-term state and local government debt. Although money market funds have been a substitute for bank savings accounts since the early 1980s, they are not insured in the same way as bank deposits.

8.2 Bond Mutual Funds

A bond mutual fund is an investment fund consisting of a portfolio of individual bonds and, occasionally, preferred shares. The net asset value of the fund is the sum of the value of each bond in the portfolio divided by the number of shares. Investors in the mutual fund hold shares, which account for their pro-rata share or interest in the portfolio. The major difference between a bond mutual fund and a money market fund is the maturity of the underlying assets. In a money market fund the maturity is as short as overnight and rarely longer than 90 days. A bond mutual fund, however, holds bonds with maturities as short as one year and as long as 30 years (or more). Exhibit 22 illustrates the general categories of bond mutual funds.[16]

Exhibit 22 Bond Mutual Funds

Type of Bond Mutual Fund	Securities Held
Global	Domestic and non-domestic government, corporate, and securitized debt
Government	Government bonds and other government-affiliated bonds
Corporate	Corporate debt
High yield	Below investment-grade corporate debt
Inflation protected	Inflation-protected government debt
National tax-free bonds	National tax-free bonds (e.g., US municipal bonds)

8.3 Stock Mutual Funds

Historically, the largest types of mutual funds based on market value of assets under management are stock (equity) funds.

There are two types of stock mutual funds. The first is an actively managed fund in which the portfolio manager seeks outstanding performance through the selection of the appropriate stocks to be included in the portfolio. Passive management is followed by index funds that are very different from actively managed funds. Their goal is to match or track the performance of different indexes. The first index fund was introduced in 1976 by the Vanguard Group.

There are several major differences between actively managed funds and index funds. First, management fees for actively managed funds are higher than for index funds. The higher fees for actively managed funds reflect its goal to outperform an index, whereas the index fund simply aims to match the return on the index. Higher fees are required to pay for the research conducted to actively select securities. A second difference is that the level of trading in an actively managed fund is much higher than in an index fund, which depending on the jurisdiction, has tax implications. Mutual funds are often required to distribute all income and capital gains realized in the portfolio, so the actively managed fund tends to have more opportunity to realize capital gains. This results in higher taxes relative to an index fund, which uses a buy-and-hold strategy. Consequently, there is less buying and selling in an index fund and less likelihood of realizing capital gains distributions.

16 In the United States, judicial rulings on federal powers of taxation have created a distinction between (federally) taxable and (federally) tax-exempt bonds and a parallel distinction for US bond mutual funds.

8.4 Hybrid/Balanced Funds

Hybrid or balanced funds are mutual funds that invest in both bonds and stocks. These types of funds represent a small fraction of the total investment in US mutual funds but are more common in Europe. These types of funds, however, have gained popularity with the growth of lifecycle funds. Lifecycle or Target Date funds manage the asset mix based on a desired retirement date. For example, if an investor is 40 years old in 2019 and planned to retire at the age of 67, he could invest in a mutual fund with a target date of 2046 and the fund would manage the appropriate asset mix over the next 27 years. In 2019 it might be 90 percent invested in shares and 10 percent in bonds. As time passes, however, the fund would gradually change the mix of shares and bonds to reflect the appropriate mix given the time to retirement.

9 POOLED INTEREST - OTHER INVESTMENT PRODUCTS

f describe mutual funds and compare them with other pooled investment products.

A fund management service for institutions or individual investors with substantial assets is the **separately managed account** (SMA), which is also commonly referred to as a “managed account,” “wrap account,” or “individually managed account.”

SMAs are managed exclusively for the benefit of a single individual or institution. Unlike a mutual fund, the assets of an SMA are owned directly by the individual or institution. The main disadvantage of an SMA is that the required minimum investment is usually much higher than with a mutual fund.

Large institutional investors are generally the dominant users of SMAs. SMAs enable asset managers to implement an investment strategy that matches an investor’s specific objectives, portfolio constraints, and tax considerations, where applicable. For example, a public pension plan investing in an asset manager’s large value equity strategy might have a socially responsible investment preference. In this case, the plan sponsor may wish to exclude certain industries, such as tobacco and defense, while also including additional companies that are deemed favorable according to other environmental, social, and governance (ESG) considerations.

9.1 Exchange-Traded Funds

Exchange-traded funds (ETFs) are investment funds that trade on exchanges (similar to individual stocks) and are generally structured as open-end funds. ETFs represent one of the fastest-growing investment products in the asset management industry. According to BlackRock, global ETF assets increased from US$428 billion in 2005 to US$4.9 trillion as of June 2018. Long-term investors—both institutional and retail—use ETFs in building a diversified asset allocation. While ETFs are structured similarly to open-end mutual funds, some key differences exist between the two products. One difference relates to transaction price. Because they are traded on exchanges, ETFs can be transacted (and are priced) intraday. That is, ETF investors buy the shares from other investors just as if they were buying or selling shares of stock. ETF investors can also short shares or purchase the shares on margin. In contrast, mutual funds typically can be purchased or sold only once a day, and short sales or purchasing shares on margin is not allowed. Mutual fund investors buy the fund shares directly from the fund, and all investments are settled at the net asset value. In practice, the market price of the ETF is likely to be close to the net asset value of the underlying investments.

Other key differences between ETFs and mutual funds relate to transaction costs and treatment of dividends and the minimum investment amount. Dividends on ETFs are paid out to the shareholders whereas mutual funds usually reinvest the dividends. Finally, the minimum required investment in ETFs is usually smaller than that of mutual funds.

9.2 Hedge Funds

Hedge funds are private investment vehicles that typically use leverage, derivatives, and long and short investment strategies. The origin of hedge funds can be traced back as far as 1949 to a fund managed by A.W. Jones & Co. It offered a strategy of a non-correlated offset to the "long-only" position typical of most portfolios. Since then, the hedge fund industry has grown considerably, with global hedge fund assets totaling US$3.3 trillion as of May 2017.

Hedge fund investment strategies are diverse and can range from specific niche strategies (e.g., long–short financial services) to global multi-strategy approaches. Consequently, hedge funds are often used by investors for portfolio diversification purposes. In general, hedge funds share a few distinguishing characteristics:

- *Short selling*: Many hedge funds implement short positions directly or synthetically using such derivatives as options, futures, and credit default swaps.
- *Absolute return seeking:* Hedge funds often seek positive returns in all market environments.
- *Leverage:* Many hedge funds use financial leverage (bank borrowing) or implicit leverage (using derivatives). The use and amount of leverage are dependent on the investment strategy being implemented.
- *Low correlation:* Some hedge funds have historically exhibited low return correlations with traditional equity and/or fixed-income asset classes.
- *Fee structures:* Hedge funds typically charge two distinct fees: a traditional asset-based management fee (AUM fee) and an incentive (or performance) fee in which the hedge fund earns a portion of the fund's realized capital gains.[17] Hedge funds have traditionally charged management fees of 2% and incentive fees of up to 20%, although there has been downward pressure on those fees amid increased competition and the availability of competing products.

Hedge funds are not readily available to all investors. They typically require a high minimum investment and often have restricted liquidity by allowing only periodic (e.g., quarterly) withdrawals or having a long fixed-term commitment.

9.3 Private Equity and Venture Capital Funds

Private equity funds and **venture capital funds** are alternative funds that seek to buy, optimize, and ultimately sell portfolio companies to generate profits. As of December 2017, assets under management in the private equity industry totaled US$3.1 trillion, a historical high point.[18] Most private equity and venture capital funds have a lifespan of approximately 7–10 years (but usually subject to contractual

17 Performance fees are often subject to high-water mark provisions, which preclude a manager from earning a performance fee unless the value of a fund at the end of a predefined measurement period is higher than the value of the fund at the beginning of the measurement period. The unpredictability of future performance leads to uncertainty in performance fee revenue, which is regarded as less reliable than revenue derived from management fees.

18 https://www.pionline.com/article/20180724/ONLINE/180729930/preqin-private-equity-aum-grows-20-in-2017-to-record-306-trillion# (accessed 13 November 2018)

extensions). Unlike most traditional asset managers that trade in public securities, private equity and venture firms often take a "hands-on" approach to their portfolio companies through a combination of financial engineering (e.g., realizing expense synergies, changing capital structures), installment of executive management and board members, and significant contributions to the development of a target company's business strategy. The final investment stage, often referred to as the "exit" or "harvesting" stage, occurs when a private equity or venture capital fund divests its portfolio companies through a merger with another company, the acquisition by another company, or an initial public offering (IPO).

As with most alternative funds, the majority of private equity and venture capital funds are structured as limited partnerships. These limited partnership agreements exist between the fund manager, called the general partner (GP), and the fund's investors, called limited partners (LPs). The funds generate revenue through several types of fees:

- *Management fees:* Fees are based on committed capital (or sometimes net asset value or invested capital) and typically range from 1–3% annually. Sometimes these fees step down several years into the investment period of a fund.
- *Transaction fees:* Fees are paid by portfolio companies to the fund for various corporate and structuring services. Typically, a percentage of the transaction fee is shared with the LPs by offsetting the management fee.
- *Carried interest:* Carried interest is the GP's share of profits (typically 20%) on sales of portfolio companies. Most GPs do not earn the incentive fee until LPs have recovered their initial investment.
- *Investment income.* Investment income includes profits generated on capital contributed to the fund by the GP.

SUMMARY

- A portfolio approach to investing could be preferable to simply investing in individual securities.
- The problem with focusing on individual securities is that this approach may lead to the investor "putting all her eggs in one basket."
- Portfolios provide important diversification benefits, allowing risk to be reduced without necessarily affecting or compromising return.
- Understanding the needs of your client and preparing an investment policy statement represent the first steps of the portfolio management process. Those steps are followed by asset allocation, security analysis, portfolio construction, portfolio monitoring and rebalancing, and performance measurement and reporting.
- Types of investors include individual and institutional investors. Institutional investors include defined benefit pension plans, endowments and foundations, banks, insurance companies, and sovereign wealth funds.
- The asset management industry is an integral component of the global financial services sector. Asset managers offer either active management, passive management, or both. Asset managers are typically categorized as traditional or alternative, although the line between traditional and alternative has blurred.

- Three key trends in the asset management industry include the growth of passive investing, "big data" in the investment process, and robo-advisers in the wealth management industry.
- Investors use different types of investment products in their portfolios. These include mutual funds, separately managed accounts, exchange-traded funds, hedge funds, and private equity and venture capital funds.

REFERENCES

Lintner, John. 1965. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics*, vol. 47, no. 1 (February):13–37.

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

Sharpe, William F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance*, vol. 19, no. 3 (September):425–442.

Singletary, Michelle. 2001. "Cautionary Tale of an Enron Employee Who Went for Broke." Seattlepi.com (10 December): http://www.seattlepi.com/money/49894_singletary10.shtml.

Treynor, J. L. 1961. "Toward a Theory of Market Value of Risky Assets." Unpublished manuscript.

PRACTICE PROBLEMS

1 Investors should use a portfolio approach to:
  A reduce risk.
  B monitor risk.
  C eliminate risk.
2 Which of the following is the *best* reason for an investor to be concerned with the composition of a portfolio?
  A Risk reduction.
  B Downside risk protection.
  C Avoidance of investment disasters.
3 With respect to the formation of portfolios, which of the following statements is *most accurate*?
  A Portfolios affect risk less than returns.
  B Portfolios affect risk more than returns.
  C Portfolios affect risk and returns equally.
4 Which of the following institutions will *on average* have the greatest need for liquidity?
  A Banks.
  B Investment companies.
  C Non-life insurance companies.
5 Which of the following institutional investors will *most likely* have the longest time horizon?
  A Defined benefit plan.
  B University endowment.
  C Life insurance company.
6 A defined benefit plan with a large number of retirees is *likely* to have a high need for:
  A income.
  B liquidity.
  C insurance.
7 Which of the following institutional investors is *most likely* to manage investments in mutual funds?
  A Insurance companies.
  B Investment companies.
  C University endowments.
8 With respect to the portfolio management process, the asset allocation is determined in the:
  A planning step.
  B feedback step.
  C execution step.
9 The planning step of the portfolio management process is *least likely* to include an assessment of the client's

A securities.

B constraints.

C risk tolerance.

10 With respect to the portfolio management process, the rebalancing of a portfolio's composition is *most likely* to occur in the:

A planning step.

B feedback step.

C execution step.

11 An analyst gathers the following information for the asset allocations of three portfolios:

Portfolio	Fixed Income (%)	Equity (%)	Alternative Assets (%)
1	25	60	15
2	60	25	15
3	15	60	25

Which of the portfolios is *most likely* appropriate for a client who has a high degree of risk tolerance?

A Portfolio 1.

B Portfolio 2.

C Portfolio 3.

12 Which of the following investment products is *most likely* to trade at their net asset value per share?

A Exchange traded funds.

B Open-end mutual funds.

C Closed-end mutual funds.

13 Which of the following financial products is *least likely* to have a capital gain distribution?

A Exchange traded funds.

B Open-end mutual funds.

C Closed-end mutual funds.

14 Which of the following forms of pooled investments is subject to the *least* amount of regulation?

A Hedge funds.

B Exchange traded funds.

C Closed-end mutual funds.

15 Which of the following pooled investments is *most likely* characterized by a few large investments?

A Hedge funds.

B Buyout funds.

C Venture capital funds.

SOLUTIONS

1 A is correct. Combining assets into a portfolio should reduce the portfolio's volatility. Specifically, "individuals and institutions should hold portfolios to reduce risk." As illustrated in the reading, however, risk reduction may not be as great during a period of dramatic economic change.

2 A is correct. Combining assets into a portfolio should reduce the portfolio's volatility. The portfolio approach does not necessarily provide downside protection or guarantee that the portfolio always will avoid losses.

3 B is correct. As illustrated in the reading, portfolios reduce risk more than they increase returns.

4 A is correct. The excess reserves invested by banks need to be relatively liquid. Although investment companies and non-life insurance companies have high liquidity needs, the liquidity need for banks is on average the greatest.

5 B is correct. Most foundations and endowments are established with the intent of having perpetual lives. Although defined benefit plans and life insurance companies have portfolios with a long time horizon, they are not perpetual.

6 A is correct. Income is necessary to meet the cash flow obligation to retirees. Although defined benefit plans have a need for income, the need for liquidity typically is quite low. A retiree may need life insurance; however, a defined benefit plan does not need insurance.

7 B is correct. Investment companies manage investments in mutual funds. Although endowments and insurance companies may own mutual funds, they do not issue or redeem shares of mutual funds.

8 C is correct. The client's objectives and constraints are established in the investment policy statement and are used to determine the client's target asset allocation, which occurs in the execution step of the portfolio management process.

9 A is correct. Securities are analyzed in the execution step. In the planning step, a client's objectives and constraints are used to develop the investment policy statement.

10 B is correct. Portfolio monitoring and rebalancing occurs in the feedback step of the portfolio management process.

11 C is correct. Portfolio 3 has the same equity exposure as Portfolio 1 and has a higher exposure to alternative assets, which have greater volatility (as discussed in the section of the reading comparing the endowments from Yale University and the University of Virginia).

12 B is correct. Open-end funds trade at their net asset value per share, whereas closed-end funds and exchange traded funds can trade at a premium or a discount.

13 A is correct. Exchange traded funds do not have capital gain distributions. If an investor sells shares of an ETF (or open-end mutual fund or closed-end mutual fund), the investor may have a capital gain or loss on the shares sold; however, the gain (or loss) from the sale is not a distribution.

14 A is correct. Hedge funds are currently exempt from the reporting requirements of a typical public investment company.

15 B is correct. Buyout funds or private equity firms make only a few large investments in private companies with the intent of selling the restructured companies in three to five years. Venture capital funds also have a short time horizon; however, these funds consist of many small investments in companies with the expectation that only a few will have a large payoff (and that most will fail).

READING

49

Portfolio Risk and Return: Part I

by Vijay Singal, PhD, CFA

Vijay Singal, PhD, CFA, is at Virginia Tech (USA).

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**a.** calculate and interpret major return measures and describe their appropriate uses;
☐	**b.** compare the money-weighted and time-weighted rates of return and evaluate the performance of portfolios based on these measures;
☐	**c.** describe characteristics of the major asset classes that investors consider in forming portfolios;
☐	**d.** calculate and interpret the mean, variance, and covariance (or correlation) of asset returns based on historical data;
☐	**e.** explain risk aversion and its implications for portfolio selection;
☐	**f.** calculate and interpret portfolio standard deviation;
☐	**g.** describe the effect on a portfolio's risk of investing in assets that are less than perfectly correlated;
☐	**h.** describe and interpret the minimum-variance and efficient frontiers of risky assets and the global minimum-variance portfolio;
☐	**i.** explain the selection of an optimal portfolio, given an investor's utility (or risk aversion) and the capital allocation line.

INTRODUCTION

1

Construction of an optimal portfolio is an important objective for an investor. In this reading, we will explore the process of examining the risk and return characteristics of individual assets, creating all possible portfolios, selecting the most efficient portfolios, and ultimately choosing the optimal portfolio tailored to the individual in question.

During the process of constructing the optimal portfolio, several factors and investment characteristics are considered. The most important of those factors are risk and return of the individual assets under consideration. Correlations among individual assets along with risk and return are important determinants of portfolio risk. Creating a portfolio for an investor requires an understanding of the risk profile of the investor. Although we will not discuss the process of determining risk aversion for individuals or institutional investors, it is necessary to obtain such information for making an informed decision. In this reading, we will explain the broad types of investors and how their risk–return preferences can be formalized to select the optimal portfolio from among the infinite portfolios contained in the investment opportunity set.

The reading is organized as follows: Sections 2–8 discuss the investment characteristics of assets. In particular, we show the various types of returns and risks, their computation and their applicability to the selection of appropriate assets for inclusion in a portfolio. Sections 9–11 discuss risk aversion and how indifference curves, which incorporate individual preferences, can be constructed. The indifference curves are then applied to the selection of an optimal portfolio using two risky assets. Sections 12–14 provides an understanding and computation of portfolio risk. The role of correlation and diversification of portfolio risk are examined in detail. Sections 15–17 begins with the risky assets available to investors and constructs a large number of risky portfolios. It illustrates the process of narrowing the choices to an efficient set of risky portfolios before identifying the optimal risky portfolio. The risky portfolio is combined with investor risk preferences to generate the investor's optimal portfolio. A summary concludes this reading.

2 INVESTMENT CHARACTERISTICS OF ASSETS: RETURN

a calculate and interpret major return measures and describe their appropriate uses;

Financial assets are frequently defined in terms of their risk and return characteristics. Comparison along these two dimensions simplifies the process of building a portfolio from among the multitude of available assets. In this section, we will compute, evaluate, and compare various measures of return and risk.

2.1 Return

Financial assets normally generate two types of return for investors. First, they may provide periodic income through cash dividends or interest payments. Second, the price of a financial asset can increase or decrease, leading to a capital gain or loss.

Some financial assets provide return through only one of these mechanisms. For example, investors in non-dividend-paying stocks obtain their return from price movement only. Similarly, you could also own or have a claim to assets that only generate periodic income. For example, defined benefit pension plans and retirement annuities make income payments as long as you live.

In the following section, we consider the computation and application of various types of returns.

2.1.1 *Holding Period Return*

Returns can be measured over a single period or over multiple periods. Single period returns are straightforward because there is only one way to calculate them. Multiple period returns, however, can be calculated in various ways and it is important to be aware of these differences to avoid confusion.

A **holding period return** is the return earned from holding an asset for a single specified period of time. The period may be 1 day, 1 week, 1 month, 5 years, or any specified period. If the asset (bond, stock, etc.) is bought now, time (t = 0), at a price of 100 and sold later, say at time (t = 1), at a price of 105 with no dividends or other income, then the holding period return is 5 percent [(105 – 100)/100)]. If the asset also pays an income of 2 units at time (t = 1), then the total return is 7 percent. This return can be generalized and shown as a mathematical expression in which P is the price and I is the income:

$$R = \frac{\left(P_1 - P_0\right) + I_1}{P_0}$$

The subscript indicates the time of the price or income, (t = 0), is the beginning of the period and (t = 1) is the end of the period. The following two observations are important.

- We computed a capital gain of 5 percent and a dividend yield of 2 percent in the above example. For ease of illustration, we assumed that the dividend is paid at time t = 1. If the dividend was received any time before t = 1, our holding period return would have been higher because we would have earned a return by reinvesting the dividend for the remainder of the period.
- Return can be expressed in decimals (0.07), fractions (7/100), or as a percent (7%). They are all equivalent.

The holding period return can be computed for a period longer than one year. For example, you may need to compute a three-year holding period return from three annual returns. In that case, the holding period return is computed by compounding the three annual returns: $R = [(1 + R_1) \times (1 + R_2) \times (1 + R_3)] - 1$, where R_1, R_2, and R_3 are the three annual returns.

2.1.2 *Arithmetic or Mean Return*

When assets have returns for multiple holding periods, it is necessary to aggregate those returns into one overall return for ease of comparison and understanding. Most holding period returns are reported as daily, monthly, or annual returns. When comparing returns across assets, it is important that the returns are computed using a common time period.

There are different methods for aggregating returns across several holding periods. The remainder of this section presents various ways of computing average returns and discusses their applicability.

The simplest way to compute the return is to take a simple arithmetic average of all holding period returns. Thus, three annual returns of –50 percent, 35 percent, and 27 percent will give us an average of 4 percent per year $= \left(\frac{-50\% + 35\% + 27\%}{3}\right)$. The arithmetic average return is easy to compute and has known statistical properties, such as standard deviation. We can calculate its standard deviation to determine how dispersed the observations are around the mean or if the mean return is statistically different from zero.

In general, the arithmetic or mean return is denoted by $\overline{R}_i$ and given by the following equation for asset i, where R_{it} is the return in period t and T is the total number of periods:

$$\overline{R}_i = \frac{R_{i1} + R_{i2} + \ldots + R_{i,T-1} + R_{iT}}{T} = \frac{1}{T}\sum_{t=1}^{T} R_{it}$$

2.1.3 *Geometric Mean Return*

The arithmetic mean return assumes that the amount invested at the beginning of each period is the same. In an investment portfolio, however, even if there are no cash flows into or out of the portfolio, the base amount changes each year. (The previous year's earnings must be added to the beginning value of the subsequent year's investment—these earnings will be "compounded" by the returns earned in that subsequent year.) We can use the geometric mean return to account for the compounding of returns.

A geometric mean return provides a more accurate representation of the growth in portfolio value over a given time period than does an arithmetic mean return. In general, the geometric mean return is denoted by $\overline{R}_{Gi}$ and given by the following equation for asset i:

$$\overline{R}_{Gi} = \sqrt[T]{\left(1 + R_{i1}\right) \times \left(1 + R_{i2}\right) \times \ldots \times \left(1 + R_{i,T-1}\right) \times \left(1 + R_{iT}\right)} - 1$$

where R_{it} is the return in period t and T is the total number of periods.

In the example in Section 2, we calculated the arithmetic mean to be 4 percent. Exhibit 1 shows the actual return for each year and the balance at the end of each year using actual returns. Beginning with an initial investment of €1.0000, we will have a balance of €0.8573 at the end of the three-year period as shown in the third column. Note that we compounded the returns because, unless otherwise stated, we earn a return on the balance as of the end of the prior year. That is, we will receive a return of 35 percent in the second year on the balance at the end of the first year, which is only €0.5000, not the initial balance of €1.0000. Let us compare the balance at the end of the three-year period computed using geometric returns with the balance we would calculate using the 4 percent annual arithmetic mean return from our earlier example. The ending value using the arithmetic mean return is €1.1249 (=1.0000 × 1.04^3). This is much larger than the actual balance of €0.8573. In general, the arithmetic return is biased upward unless each of the underlying holding period returns are equal. The bias in arithmetic mean returns is particularly severe if holding period returns are a mix of both positive and negative returns, as in the example.

Exhibit 1

	Actual Return for the Year (%)	Year-End Amount	Year-End Amount Using Arithmetic Return of 4%	Year-End Amount Using Geometric Return of –5%
Year 0		€1.0000	€1.0000	€1.0000
Year 1	–50	0.5000	1.0400	0.9500
Year 2	35	0.6750	1.0816	0.9025
Year 3	27	0.8573	1.1249	0.8574

MONEY-WEIGHTED RETURN OR INTERNAL RATE OF RETURN

3

b compare the money-weighted and time-weighted rates of return and evaluate the performance of portfolios based on these measures;

The arithmetic and geometric return computations do not account for the cash flows into and out of a portfolio If the investor had invested €10,000 in the first year, €1,000 in the second year, and €1,000 in the third year, then the return of −50 percent in the first year significantly hurts her. On the other hand, if she had invested only €100 in the first year, the effect of the −50 percent return is drastically reduced.

The **money-weighted return** accounts for the money invested and provides the investor with information on the return she earns on her actual investment. The money-weighted return and its calculation are similar to the **internal rate of return** and the yield to maturity. Just like the internal rate of return, amounts invested are cash outflows from the investor's perspective and amounts returned or withdrawn by the investor, or the money that remains at the end of an investment cycle, is a cash inflow for the investor.

The money-weighted return can be illustrated most effectively with an example. In this example, we use the returns from the previous example. Assume that the investor invests €100 in a mutual fund at the beginning of the first year, adds another €950 at the beginning of the second year, and withdraws €350 at the end of the second year. The cash flows are shown in Exhibit 2.

Exhibit 2

Year	1	2	3
Balance from previous year	€0	€50	€1,000
New investment by the investor (cash inflow for the mutual fund) at the start of the year	100	950	0
Net balance at the beginning of year	100	1,000	1,000
Investment return for the year	−50%	35%	27%
Investment gain (loss)	−50	350	270
Withdrawal by the investor (cash outflow for the mutual fund) at the end of the year	0	−350	0
Balance at the end of year	€50	€1,000	€1,270

The internal rate of return is the discount rate at which the sum of present values of these cash flows will equal zero. In general, the equation may be expressed as follows, where T is the number of periods, CF_t is the cash flow at time t, and IRR is the internal rate of return or the money-weighted rate of return:

$$\sum_{t=0}^{T} \frac{CF_t}{(1 + \text{IRR})^t} = 0$$

A cash flow can be positive or negative; a positive cash flow is an inflow where money flows to the investor, whereas a negative cash flow is an outflow where money flows away from the investor. We can compute the internal rate of return by using the above equation. The flows are expressed as follows, where each cash inflow or outflow occurs at the end of each year. Thus, CF_0 refers to the cash flow at the end of Year 0 or beginning of Year 1, and CF_3 refers to the cash flow at end of Year 3 or beginning of Year 4. Because cash flows are being discounted to the present—that is, end of Year 0 or beginning of Year 1—the period of discounting CF_0 is zero.

$$CF_0 = -100$$
$$CF_1 = -950$$
$$CF_2 = +350$$
$$CF_3 = +1{,}270$$
$$\frac{CF_0}{(1+IRR)^0} + \frac{CF_1}{(1+IRR)^1} + \frac{CF_2}{(1+IRR)^2} + \frac{CF_3}{(1+IRR)^3}$$
$$= \frac{-100}{1} + \frac{-950}{(1+IRR)^1} + \frac{+350}{(1+IRR)^2} + \frac{+1270}{(1+IRR)^3} = 0$$
$$IRR = 26.11\%$$

IRR = 26.11% is the internal rate of return, or the money-weighted rate of return, which tells the investor what she earned on the actual euros invested for the entire period. This return is much greater than the arithmetic and geometric mean returns because only a small amount was invested when the mutual fund's return was −50 percent.

Next, we'll illustrate calculating the money-weighted return for a dividend paying stock. Consider an investment that covers a two-year horizon. At time $t = 0$, an investor buys one share at \$200. At time $t = 1$, he purchases an additional share at \$225. At the end of Year 2, $t = 2$, he sells both shares for \$235 each. During both years, the stock pays a per-share dividend of \$5. The $t = 1$ dividend is not reinvested. Exhibit 3 shows the total cash inflows and outflows.

Exhibit 3 Cash Flows

Time	Outflows
0	\$200 to purchase the first share
1	\$225 to purchase the second share

Time	Inflows
1	\$5 dividend received from first share (and not reinvested)
2	\$10 dividend (\$5 per share × 2 shares) received
2	\$470 received from selling two shares at \$235 per share

To solve for the money-weighted return, we use either a financial calculator that allows us to enter cash flows or a spreadsheet with an IRR function.[1] The first step is to group net cash flows by time. For this example, we have −\$200 for the $t = 0$ net

1 In this particular case we could solve for r by solving the quadratic equation $480x^2 - 220x - 200 = 0$ with $x = 1/(1 + r)$, using standard results from algebra. In general, however, we rely on a calculator or spreadsheet software to compute a money-weighted rate of return.

cash flow, −\$220 = −\$225 + \$5 for the t = 1 net cash flow, and \$480 for the t = 2 net cash flow. After entering these cash flows, we use the spreadsheet's or calculator's IRR function to find that the money-weighted rate of return is 9.39 percent.[2]

$$CF_0 = -200$$
$$CF_1 = -220$$
$$CF_2 = +480$$
$$\frac{CF_0}{(1+IRR)^0} + \frac{CF_1}{(1+IRR)^1} + \frac{CF_2}{(1+IRR)^2}$$
$$= \frac{-200}{1} + \frac{-220}{(1+IRR)^1} + \frac{480}{(1+IRR)^2} = 0$$
$$IRR = 9.39\%$$

Now we take a closer look at what has happened to the portfolio during each of the two years. In the first year, the portfolio generated a one-period holding period return of (\$5 + \$225 − \$200)/\$200 = 15 percent. At the beginning of the second year, the amount invested is \$450, calculated as \$225 (per share price of stock) × 2 shares, because the \$5 dividend was spent rather than reinvested. At the end of the second year, the proceeds from the liquidation of the portfolio are \$470 (as detailed in Exhibit 3) plus \$10 in dividends (as also detailed in Exhibit 3). So in the second year the portfolio produced a holding period return of (\$10 + \$470 − \$450)/\$450 = 6.67 percent. The mean holding period return was (15% + 6.67%)/2 = 10.84 percent. The money-weighted rate of return, which we calculated as 9.39 percent, puts a greater weight on the second year's relatively poor performance (6.67 percent) than the first year's relatively good performance (15 percent), as more money was invested in the second year than in the first. That is the sense in which returns in this method of calculating performance are "money weighted." Although the money-weighted return is an accurate measure of what the investor actually earned on the money invested, it is limited in its applicability to other situations. For example, it does not allow for return comparison between different individuals or different investment opportunities. Two investors in the *same* mutual fund or with the same portfolio of underlying investments may have different money-weighted returns because they invested different amounts in different years.

EXAMPLE 1

Computation of Returns

Ulli Lohrmann and his wife, Suzanne Lohrmann, are planning for retirement and want to compare the past performance of a few mutual funds they are considering for investment. They believe that a comparison over a five-year period would be appropriate. They are given the following information about the Rhein Valley Superior Fund that they are considering.

Year	Assets Under Management at the Beginning of Year (€)	Net Return (%)
1	30 million	15
2	45 million	−5
3	20 million	10

(continued)

2 Note that the calculator or spreadsheet will give the IRR as a periodic rate. If the periods are not annual, we annualize the periodic rate.

Year	Assets Under Management at the Beginning of Year (€)	Net Return (%)
4	25 million	15
5	35 million	3

The Lohrmanns are interested in aggregating this information for ease of comparison with other funds.

1 Compute the holding period return for the five-year period.
2 Compute the arithmetic mean annual return.
3 Compute the geometric mean annual return. How does it compare with the arithmetic mean annual return?
4 The Lohrmanns want to earn a minimum annual return of 5 percent. Is the money-weighted annual return greater than 5 percent?

Solution to 1:

The holding period return is $R = (1 + R_1)(1 + R_2)(1 + R_3)(1 + R_4)(1 + R_5) - 1 = (1.15)(0.95)(1.10)(1.15)(1.03) - 1 = 0.4235 = 42.35\%$ for the five-year period.

Solution to 2:

The arithmetic mean annual return can be computed as an arithmetic mean of the returns given by this equation:

$$\overline{R}_i = \frac{15\% - 5\% + 10\% + 15\% + 3\%}{5} = 7.60\%$$

Solution to 3:

The geometric mean annual return can be computed using this equation:

$$\begin{aligned}\overline{R}_{Gi} &= \sqrt[T]{\left(1 + R_{i1}\right) \times \left(1 + R_{i2}\right) \times \ldots \times \left(1 + R_{i,T-1}\right) \times \left(1 + R_{iT}\right)} - 1 \\ &= \sqrt[5]{1.15 \times 0.95 \times 1.10 \times 1.15 \times 1.03} - 1 \\ &= \sqrt[5]{1.4235} - 1 = 0.0732 = 7.32\%\end{aligned}$$

Thus, the geometric mean annual return is 7.32 percent, slightly less than the arithmetic mean return.

Solution to 4:

To calculate the money-weighted rate of return, tabulate the annual returns and investment amounts to determine the cash flows, as shown in Exhibit 4. All amounts are in millions of euros.

Exhibit 4

Year	1	2	3	4	5
Balance from previous year	0	34.50	42.75	22.00	28.75
New investment by the investor (cash inflow for the Rhein fund)	30.00	10.50	0	3.00	6.25
Withdrawal by the investor (cash outflow for the Rhein fund)	0	0	–22.75	0	0

Exhibit 4 (Continued)

Year	1	2	3	4	5
Net balance at the beginning of year	30.00	45.00	20.00	25.00	35.00
Investment return for the year	15%	−5%	10%	15%	3%
Investment gain (loss)	4.50	−2.25	2.00	3.75	1.05
Balance at the end of year	34.50	42.75	22.00	28.75	36.05

CF_0 = −30.00, CF_1 = −10.50, CF_2 = +22.75, CF_3 = −3.00, CF_4 = −6.25, CF_5 = +36.05.

For clarification, it may be appropriate to explain the notation for cash flows. Each cash inflow or outflow occurs at the end of each year. Thus, CF_0 refers to the cash flow at the end of Year 0 or beginning of Year 1, and CF_5 refers to the cash flow at end of Year 5 or beginning of Year 6. Because cash flows are being discounted to the present—that is, end of Year 0 or beginning of Year 1—the period of discounting CF_0 is zero whereas the period of discounting for CF_5 is 5 years.

To get the exact money-weighted rate of return (IRR), the following equation would be equal to zero. Instead of calculating, however, use the 5 percent return to see whether the value of the expression is positive or not. If it is positive, then the money-weighted rate of return is greater than 5 percent, because a 5 percent discount rate could not reduce the value to zero.

$$\frac{-30.00}{(1.05)^0} + \frac{-10.50}{(1.05)^1} + \frac{22.75}{(1.05)^2} + \frac{-3.00}{(1.05)^3} + \frac{-6.25}{(1.05)^4} + \frac{36.05}{(1.05)^5} = 1.1471$$

Because the value is positive, the money-weighted rate of return is greater than 5 percent. Using a financial calculator, the exact money-weighted rate of return is 5.86 percent.

TIME-WEIGHTED RATE OF RETURN

4

b compare the money-weighted and time-weighted rates of return and evaluate the performance of portfolios based on these measures;

An investment measure that is not sensitive to the additions and withdrawals of funds is the time-weighted rate of return. The **time-weighted rate of return** measures the compound rate of growth of $1 initially invested in the portfolio over a stated measurement period. For the evaluation of portfolios of publicly traded securities, the time-weighted rate of return is the preferred performance measure as it neutralizes the effect of cash withdrawals or additions to the portfolio, which are generally outside of the control of the portfolio manager. To compute an exact time-weighted rate of return on a portfolio, take the following three steps:

1 Price the portfolio immediately prior to any significant addition or withdrawal of funds. Break the overall evaluation period into subperiods based on the dates of cash inflows and outflows.

2 Calculate the holding period return on the portfolio for each subperiod.
3 Link or compound holding period returns to obtain an annual rate of return for the year (the time-weighted rate of return for the year). If the investment is for more than one year, take the geometric mean of the annual returns to obtain the time-weighted rate of return over that measurement period.

Let us return to our dividend stock money-weighted example and calculate the time-weighted rate of return for that investor's portfolio. In that example, we computed the holding period returns on the portfolio, Step 2 in the procedure for finding the time-weighted rate of return. Given that the portfolio earned returns of 15 percent during the first year and 6.67 percent during the second year, what is the portfolio's time-weighted rate of return over an evaluation period of two years?

We find this time-weighted return by taking the geometric mean of the two holding period returns, Step 3 in the procedure above. The calculation of the geometric mean exactly mirrors the calculation of a compound growth rate. Here, we take the product of 1 plus the holding period return for each period to find the terminal value at t = 2 of \$1 invested at t = 0. We then take the square root of this product and subtract 1 to get the geometric mean. We interpret the result as the annual compound growth rate of \$1 invested in the portfolio at t = 0. Thus we have

$$(1 + \text{Time-weighted return})^2 = (1.15)(1.0667)$$

$$\text{Time-weighted return} = \sqrt{(1.15)(1.0667)} - 1 = 10.76\%$$

The time-weighted return on the portfolio was 10.76 percent, compared with the money-weighted return of 9.39 percent, which gave larger weight to the second year's return. We can see why investment managers find time-weighted returns more meaningful. If a client gives an investment manager more funds to invest at an unfavorable time, the manager's money-weighted rate of return will tend to be depressed. If a client adds funds at a favorable time, the money-weighted return will tend to be elevated. The time-weighted rate of return removes these effects.

In defining the steps to calculate an exact time-weighted rate of return, we said that the portfolio should be valued immediately prior to any significant addition or withdrawal of funds. With the amount of cash flow activity in many portfolios, this task can be costly. We can often obtain a reasonable approximation of the time-weighted rate of return by valuing the portfolio at frequent, regular intervals, particularly if additions and withdrawals are unrelated to market movements. The more frequent the valuation, the more accurate the approximation. Daily valuation is commonplace. Suppose that a portfolio is valued daily over the course of a year. To compute the time-weighted return for the year, we first compute each day's holding period return. We compute 365 such daily returns, denoted $r_1, r_2, \ldots, r_{365}$. We obtain the annual return for the year by linking the daily holding period returns in the following way: $(1 + r_1) \times (1 + r_2) \times \ldots \times (1 + r_{365}) - 1$. If withdrawals and additions to the portfolio happen only at day's end, this annual return is a precise time-weighted rate of return for the year. Otherwise, it is an approximate time-weighted return for the year.

If we have a number of years of data, we can calculate a time-weighted return for each year individually, as above. If r_i is the time-weighted return for year i, we calculate an annualized time-weighted return as the geometric mean of N annual returns, as follows:

$$r_{TW} = \left[(1 + r_1) \times (1 + r_2) \times \ldots \times (1 + r_N)\right]^{1/N} - 1$$

Example 2 illustrates the calculation of the time-weighted rate of return.

EXAMPLE 2

Time-Weighted Rate of Return

Strubeck Corporation sponsors a pension plan for its employees. It manages part of the equity portfolio in-house and delegates management of the balance to Super Trust Company. As chief investment officer of Strubeck, you want to review the performance of the in-house and Super Trust portfolios over the last four quarters. You have arranged for outflows and inflows to the portfolios to be made at the very beginning of the quarter. Exhibit 5 summarizes the inflows and outflows as well as the two portfolios' valuations. In the table, the ending value is the portfolio's value just prior to the cash inflow or outflow at the beginning of the quarter. The amount invested is the amount each portfolio manager is responsible for investing.

Exhibit 5 Cash Flows for the In-House Strubeck Account and the Super Trust Account

	Quarter			
	1 ($)	2 ($)	3 ($)	4 ($)
In-House Account				
Beginning value	4,000,000	6,000,000	5,775,000	6,720,000
Beginning of period inflow (outflow)	1,000,000	(500,000)	225,000	(600,000)
Amount invested	5,000,000	5,500,000	6,000,000	6,120,000
Ending value	6,000,000	5,775,000	6,720,000	5,508,000
Super Trust Account				
Beginning value	10,000,000	13,200,000	12,240,000	5,659,200
Beginning of period inflow (outflow)	2,000,000	(1,200,000)	(7,000,000)	(400,000)
Amount invested	12,000,000	12,000,000	5,240,000	5,259,200
Ending value	13,200,000	12,240,000	5,659,200	5,469,568

Based on the information given, address the following.

1 Calculate the time-weighted rate of return for the in-house account.

2 Calculate the time-weighted rate of return for the Super Trust account.

Solution to 1:

To calculate the time-weighted rate of return for the in-house account, we compute the quarterly holding period returns for the account and link them into an annual return. The in-house account's time-weighted rate of return is 27 percent, calculated as follows:

$$\text{1Q HPR: } r_1 = (\$6{,}000{,}000 - \$5{,}000{,}000)/\$5{,}000{,}000 = 0.20$$

$$\text{2Q HPR: } r_2 = (\$5{,}775{,}000 - \$5{,}500{,}000)/\$5{,}500{,}000 = 0.05$$

$$\text{3Q HPR: } r_3 = (\$6{,}720{,}000 - \$6{,}000{,}000)/\$6{,}000{,}000 = 0.12$$

$$\text{4Q HPR: } r_4 = (\$5{,}508{,}000 - \$6{,}120{,}000)/\$6{,}120{,}000 = -0.10$$

$$(1 + r_1)(1 + r_2)(1 + r_3)(1 + r_4) - 1 = (1.20)(1.05)(1.12)(0.90) - 1 = 0.27 \text{ or } 27\%$$

Solution to 2:

The account managed by Super Trust has a time-weighted rate of return of 26 percent, calculated as follows:

$$\text{1Q HPR: } r_1 = (\$13{,}200{,}000 - \$12{,}000{,}000)/\$12{,}000{,}000 = 0.10$$

$$\text{2Q HPR: } r_2 = (\$12{,}240{,}000 - \$12{,}000{,}000)/\$12{,}000{,}000 = 0.02$$

$$\text{3Q HPR: } r_3 = (\$5{,}659{,}200 - \$5{,}240{,}000)/\$5{,}240{,}000 = 0.08$$

$$\text{4Q HPR: } r_4 = (\$5{,}469{,}568 - \$5{,}259{,}200)/\$5{,}259{,}200 = 0.04$$

$$(1 + r_1)(1 + r_2)(1 + r_3)(1 + r_4) - 1 = (1.10)(1.02)(1.08)(1.04) - 1 = 0.26 \text{ or } 26\%$$

The in-house portfolio's time-weighted rate of return is higher than the Super Trust portfolio's by 100 basis points.

Having worked through this exercise, we are ready to look at a more detailed case.

EXAMPLE 3

Time-Weighted and Money-Weighted Rates of Return Side by Side

Your task is to compute the investment performance of the Walbright Fund during 2014. The facts are as follows:

- On 1 January 2014, the Walbright Fund had a market value of $100 million.
- During the period 1 January 2014 to 30 April 2014, the stocks in the fund showed a capital gain of $10 million.
- On 1 May 2014, the stocks in the fund paid a total dividend of $2 million. All dividends were reinvested in additional shares.
- Because the fund's performance had been exceptional, institutions invested an additional $20 million in Walbright on 1 May 2014, raising assets under management to $132 million ($100 + $10 + $2 + $20).
- On 31 December 2014, Walbright received total dividends of $2.64 million. The fund's market value on 31 December 2014, not including the $2.64 million in dividends, was $140 million.
- The fund made no other interim cash payments during 2014.

Based on the information given, address the following.

1. Compute the Walbright Fund's time-weighted rate of return.
2. Compute the Walbright Fund's money-weighted rate of return.
3. Interpret the differences between the time-weighted and money-weighted rates of return.

Solution to 1:

Because interim cash flows were made on 1 May 2014, we must compute two interim total returns and then link them to obtain an annual return. Exhibit 6 lists the relevant market values on 1 January, 1 May, and 31 December as well as the associated interim four-month (1 January to 1 May) and eight-month (1 May to 31 December) holding period returns.

Exhibit 6 Cash Flows for the Walbright Fund

1 January 2014	Beginning portfolio value = \$100 million
1 May 2014	Dividends received before additional investment = \$2 million
	Ending portfolio value = \$110 million
	Holding period return $= \dfrac{\$2 + \$10}{\$100} = 12\%$
	New investment = \$20 million
	Beginning market value for last 2/3 of year = \$132 million
31 December 2014	Dividends received = \$2.64 million
	Ending portfolio value = \$140 million
	Holding period return $= \dfrac{\$2.64 + \$140 - \$132}{\$132} = 8.06\%$

Now we must geometrically link the four- and eight-month returns to compute an annual return. We compute the time-weighted return as follows:

Time-weighted return = 1.12 × 1.0806 − 1 = 0.2103

In this instance, we compute a time-weighted rate of return of 21.03 percent for one year. The four-month and eight-month intervals combine to equal one year. (Taking the square root of the product 1.12 × 1.0806 would be appropriate only if 1.12 and 1.0806 each applied to one full year.)

Solution to 2:

To calculate the money-weighted return, we find the discount rate that sets the sum of the present value of cash inflows and outflows equal to zero. The initial market value of the fund and all additions to it are treated as cash outflows. (Think of them as expenditures.) Withdrawals, receipts, and the ending market value of the fund are counted as inflows. (The ending market value is the amount investors receive on liquidating the fund.) Because interim cash flows have occurred at four-month intervals, we must solve for the four-month internal rate of return. Exhibit 6 details the cash flows and their timing.

$CF_0 = -100$

$CF_1 = -20$

$CF_2 = 0$

$CF_3 = 142.64$

CF_0 refers to the initial investment of \$100 million made at the beginning of the first four-month interval on 1 January 2014. CF_1 refers to the cash flows made at end of the first four-month interval or the beginning of the second four-month interval on 1 May 2014. Those cash flows include a cash inflow of \$2 million for the dividend received and cash outflows of \$22 million for the dividend reinvested and additional investment respectively. The second four-month interval had no cash flow so CF_2 is equal to zero. CF_3 refers to the cash inflows at the end of the third four-month interval. Those cash inflows include a \$2.64 million dividend received and the fund's terminal market value of \$140 million

Using a spreadsheet or IRR-enabled calculator, we use −100, −20, 0, and \$142.64 for the $t = 0$, $t = 1$, $t = 2$, and $t = 3$ net cash flows, respectively.[3] Using either tool, we get a four-month IRR of 6.28 percent. The quick way to annualize this is to multiply by 3. A more accurate way is $(1.0628)^3 - 1 = 0.20$ or 20 percent.

$$\frac{CF_0}{(1+\text{IRR})^0} + \frac{CF_1}{(1+\text{IRR})^1} + \frac{CF_2}{(1+\text{IRR})^2} + \frac{CF_3}{(1+\text{IRR})^3}$$

$$= \frac{-100}{1} + \frac{-20}{(1+\text{IRR})^1} + \frac{0}{(1+\text{IRR})^2} + \frac{142.64}{(1+\text{IRR})^3}$$

$$\text{IRR} = 6.28\%$$

Solution to 3:

In this example, the time-weighted return (21.03 percent) is greater than the money-weighted return (20 percent). The Walbright Fund's performance was relatively poorer during the eight-month period, when the fund owned more shares, than it was overall. This fact is reflected in a lower money-weighted rate of return compared with time-weighted rate of return, as the money-weighted return is sensitive to the timing and amount of withdrawals and additions to the portfolio.

The accurate measurement of portfolio returns is important to the process of evaluating portfolio managers. In addition to considering returns, however, analysts must also weigh risk. When we worked through Example 2, we stopped short of suggesting that in-house management was superior to Super Trust because it earned a higher time-weighted rate of return. A judgement as to whether performance was "better" or "worse" must include the risk dimension, which will be covered later in your study materials.

5 ANNUALIZED RETURN

a calculate and interpret major return measures and describe their appropriate uses;

The period during which a return is earned or computed can vary and often we have to annualize a return that was calculated for a period that is shorter (or longer) than one year. You might buy a short-term treasury bill with a maturity of 3 months, or you might take a position in a futures contract that expires at the end of the next quarter. How can we compare these returns? In, many cases, it is most convenient to annualize all available returns. Thus, daily weekly, monthly, and quarterly returns are converted to annualize all available returns. Many formulas used for calculating certain values or prices also require all returns and periods to be expressed as annualized rates of return. For example, the most common version of the Black–Scholes option-pricing model requires annualized returns and periods to be in years.

To annualize any return for a period shorter than one year, the return for the period must be compounded by the number of periods in a year. A monthly return is compounded 12 times, a weekly return is compounded 52 times, and a quarterly return is compounded 4 times. Daily returns are normally compounded 365 times. For an uncommon number of days, we compound by the ratio of 365 to the number of days.

3 By convention, we denote outflow with a negative sign, and we need 0 as a placeholder for the $t = 2$.

If the weekly return is 0.2 percent, then the compound annual return is computed as shown because there are 52 weeks in a year:

$$r_{annual} = \left(1 + r_{weekly}\right)^{52} - 1 = \left(1 + 0.2\%\right)^{52} - 1$$
$$= \left(1.002\right)^{52} - 1 = 0.1095 = 10.95\%$$

If the return for 15 days is 0.4 percent, the annualized return is computed assuming 365 days in a year. Thus,

$$r_{annual} = \left(1 + r_{15}\right)^{365/15} - 1 = \left(1 + 0.4\%\right)^{365/15} - 1$$
$$= \left(1.004\right)^{365/15} - 1 = 0.1020 = 10.20\%$$

A general equation to annualize returns is given, where c is the number of periods in a year. For a quarter, c = 4 and for a month, c = 12:

$$r_{annual} = \left(1 + r_{period}\right)^{c} - 1$$

How can we annualize a return when the holding period return is more than one year? For example, how do we annualize an 18-month holding period return? Because one year contains two-thirds of 18-month periods, c = 2/3 in the above equation. An 18-month return of 20 percent can be annualized, as shown:

$$r_{annual} = \left(1 + r_{18month}\right)^{2/3} - 1 = \left(1 + 0.20\right)^{2/3} - 1 = 0.1292 = 12.92\%$$

Similar expressions can be constructed when quarterly or weekly returns are needed for comparison instead of annual returns. In such cases, c is equal to the number of holding periods in a quarter or in a week. For example, assume that you want to convert daily returns to weekly returns or annual returns to weekly returns for comparison between weekly returns. For converting daily returns to weekly returns, c = 5, assuming that there are five trading days in a week. For converting annual returns to weekly returns, c = 1/52. The expressions for annual returns can then be rewritten as expressions for weekly returns, as shown:

$$r_{weekly} = \left(1 + r_{daily}\right)^{5} - 1;\ r_{weekly} = \left(1 + r_{annual}\right)^{1/52} - 1$$

One major limitation of annualizing returns is the implicit assumption that returns can be repeated precisely, that is, money can be reinvested repeatedly while earning a similar return. This type of return is not always possible. An investor may earn a return of 5 percent during a week because the market went up that week or he got lucky with his stock, but it is highly unlikely that he will earn a return of 5 percent every week for the next 51 weeks, resulting in an annualized return of 1,164.3 percent (= 1.05^{52} – 1). Therefore, it is important to annualize short-term returns with this limitation in mind.

EXAMPLE 4

Annualized Returns

An analyst is trying to evaluate three securities that have been in her portfolio for different periods of time.

- In the last 100 days, Security A has earned a return of 6.2 percent.
- Security B has earned 2 percent over the last 4 weeks.
- Security C has earned a return of 5 percent over the last 3 months

The analyst is trying to assess the relative performance of the 3 securities.

Solution:

Annualized return for Security A: $R_{SA} = (1 + 0.062)^{365/100} - 1 = 0.2455 =$ 24.55%

Annualized return for Security B: $R_{SB} = (1 + 0.02)^{52/4} - 1 = 0.2936 = 29.36\%$

Annualized return for Security C: $R_{SC} = (1 + 0.05)^{4} - 1 = 0.2155 = 21.55\%$

Security B has the highest annualized return.

6 OTHER MAJOR RETURN MEASURES AND THEIR APPLICATIONS

a calculate and interpret major return measures and describe their appropriate uses;

The statistical measures of return discussed in the previous section are generally applicable across a wide range of assets and time periods. Special assets, however, such as mutual funds, and other considerations, such as taxes or inflation, may require return measures that are specific to a particular application.

Although it is not possible to consider all types of special applications, we will discuss the effect of fees (gross versus net returns), taxes (pre-tax and after-tax returns), inflation (nominal and real returns), and **leverage**. Many investors use mutual funds or other external entities (i.e., investment vehicles) for investment. In those cases, funds charge management fees and expenses to the investors. Consequently, gross and net-of-fund-expense returns should also be considered. Of course, an investor may be interested in the net-of-expenses after-tax real return, which is in fact what an investor truly receives. We consider these additional return measures in the following sections.

6.1 Gross and Net Return

A gross return is the return earned by an asset manager prior to deductions for management expenses, custodial fees, taxes, or any other expenses that are not directly related to the generation of returns but rather related to the management and administration of an investment. These expenses are not deducted from the gross return because they may vary with the amount of assets under management or may vary because of the tax status of the investor. Trading expenses, however, such as commissions, *are* accounted for in (i.e., deducted from) the computation of gross return because trading expenses contribute directly to the return earned by the manager. Thus, gross return is an appropriate measure for evaluating and comparing the investment skill of asset managers because it does not include any fees related to the management and administration of an investment.

Net return is a measure of what the investment vehicle (mutual fund, etc.) has earned for the investor. Net return accounts for (i.e., deducts) all managerial and administrative expenses that reduce an investor's return. Because individual investors are most concerned about the net return (i.e., what they actually receive), small mutual funds with a limited amount of assets under management are at a disadvantage compared with the larger funds that can spread their largely fixed administrative expenses over a larger asset base. As a result, many small mutual funds waive part of the expenses to keep the funds competitive.

6.2 Pre-tax and After-tax Nominal Return

All return measures discussed previously are pre-tax nominal returns—that is, no adjustment has been made for taxes or inflation. In general, all returns are pre-tax nominal returns unless they are otherwise designated.

Many investors are concerned about the possible tax liability associated with their returns because taxes reduce the net return that they receive. Capital gains and income may be taxed differently, depending on the jurisdiction. Capital gains come in two forms: short-term capital gains and long-term capital gains. Long-term capital gains receive preferential tax treatment in a number of countries. Interest income is taxed as ordinary income in most countries. Dividend income may be taxed as ordinary income, may have a lower tax rate, or may be exempt from taxes depending on the country and the type of investor. The after-tax nominal return is computed as the total return minus any allowance for taxes on dividends, interest and realized gains.[4]

Because taxes are paid on realized capital gains and income, the investment manager can minimize the tax liability by selecting appropriate securities (e.g., those subject to more favorable taxation, all other investment considerations equal) and reducing trading turnover. Therefore, taxable investors evaluate investment managers based on the after-tax nominal return.

6.3 Real Returns

A nominal return (r) consists of three components: a real risk-free return as compensation for postponing consumption (r_{rF}), inflation as compensation for loss of purchasing power (π), and a **risk premium** for assuming risk (RP). Thus, nominal return and real return can be expressed as:

$$(1 + r) = (1 + r_{rF}) \times (1 + \pi) \times (1 + RP)$$

$$(1 + r_{real}) = (1 + r_{rF}) \times (1 + RP) \text{ or}$$

$$(1 + r_{real}) = (1 + r) \div (1 + \pi)$$

Often the real risk-free return and the risk premium are combined to arrive at the real "risky" rate as given in the second equation above, simply referred to as the real return. Real returns are particularly useful in comparing returns across time periods because inflation rates may vary over time. Real returns are also useful in comparing returns among countries when returns are expressed in local currencies instead of a constant investor currency and when inflation rates vary between countries (which are usually the case). Finally, the after-tax real return is what the investor receives as compensation for postponing consumption and assuming risk after paying taxes on investment returns. As a result, the after-tax real return becomes a reliable benchmark for making investment decisions. Although it is a measure of an investor's benchmark return, it is not commonly calculated by asset managers because it is difficult to estimate a general tax component applicable to all investors. For example, the tax component depends on an investor's specific taxation rate (marginal tax rate), how long the investor holds an investment (long-term versus short-term), and the type of account the asset is held in (tax-exempt, tax-deferred, or normal).

4 Bonds issued at a discount to the par value may be taxed based on accrued gains instead of realized gains.

EXAMPLE 5

Computation of Special Returns

Let's return to Example 1. After reading this section, Mr. Lohrmann decided that he was not being fair to the fund manager by including the asset management fee and other expenses because the small size of the fund would put it at a competitive disadvantage. He learns that the fund spends a fixed amount of €500,000 every year on expenses that are unrelated to the manager's performance.

Mr. Lohrmann has become concerned that both taxes and inflation may reduce his return. Based on the current tax code, he expects to pay 20 percent tax on the return he earns from his investment. Historically, inflation has been around 2 percent and he expects the same rate of inflation to be maintained.

1. Estimate the annual gross return for the first year by adding back the fixed expenses.
2. What is the net return that investors in the Rhein Valley Superior Fund earned during the five-year period?
3. What is the after-tax net return for the first year that investors earned from the Rhein Valley Superior Fund? Assume that all gains are realized at the end of the year and the taxes are paid immediately at that time.
4. What is the anticipated after-tax real return that investors would have earned in the fifth year?

Solution to 1:

The gross return for the first year is higher by 1.67 percent (= €500,000/€30,000,000) than the investor return reported by the fund. Thus, the gross return is 16.67 percent (= 15% + 1.67%).

Solution to 2:

The investor return reported by the mutual fund is the net return of the fund after accounting for all direct and indirect expenses. The net return is also the pre-tax nominal return because it has not been adjusted for taxes or inflation. The net return for the five-year holding period was 42.35 percent.

Solution to 3:

The net return earned by investors during the first year was 15 percent. Applying a 20 percent tax rate, the after-tax return that accrues to the investors is 12 percent [= 15% − (0.20 × 15%)].

Solution to 4:

As in Part 3, the after-tax return earned by investors in the fifth year is 2.4 percent [= 3% − (0.20 × 3%)]. Inflation reduces the return by 2 percent so the after-tax real return earned by investors in the fifth year is 0.39 percent, as shown:

$$\frac{(1+2.40\%)}{(1+2.00\%)} - 1 = \frac{(1+0.0240)}{(1+0.0200)} - 1 = 1.0039 - 1 = 0.0039 = 0.39\%$$

Note that taxes are paid before adjusting for inflation.

6.4 Leveraged Return

In the previous calculations, we have assumed that the investor's position in an asset is equal to the total investment made by an investor using his or her own money. This section differs in that the investor creates a leveraged position. There are two ways of creating a claim on asset returns that are greater than the investment of one's own money. First, an investor may trade futures contracts in which the money required to take a position may be as little as 10 percent of the notional value of the asset. In this case, the leveraged return, the return on the investor's own money, is 10 times the actual return of the underlying security. Both the gains and losses are amplified by a factor of 10.

Investors can also invest more than their own money by borrowing money to purchase the asset. This approach is easily done in stocks and bonds, and very common when investing in real estate. If half (50 percent) of the money invested is borrowed, then the gross return to the investor is doubled but the interest to be paid on borrowed money must be deducted in order to calculate the net return.

HISTORICAL RETURN AND RISK

7

c describe characteristics of the major asset classes that investors consider in forming portfolios;

At this time, it is helpful to look at historical risk and returns for the three main asset categories: stocks, bonds, and Treasury bills. Stocks refer to corporate ownership, bonds refer to long-term fixed-income securities, and Treasury bills refer to short-term government debt securities. Although there is generally no expectation of default on government securities, long-term government bond prices are volatile (risky) because of possible future changes in interest rates. In addition, bondholders also face the risk that inflation will reduce the purchasing power of their cash flows.

7.1 Historical Mean Return and Expected Return

Before examining historical data, it is useful to distinguish between the historical mean return and expected return, which are very different concepts but easy to confuse. Historical return is what was actually earned in the *past*, whereas expected return is what an investor anticipates to earn in the *future*.

Expected return is the nominal return that would cause the marginal investor to invest in an asset based on the real risk-free interest rate (r_{rF}), expected inflation [$E(\pi)$], and expected risk premium for the risk of the asset [$E(RP)$]. The real risk-free interest rate is expected to be positive as compensation for postponing consumption. Similarly, the risk premium is expected to be positive in most cases.[5] The expected inflation rate is generally positive, except when the economy is in a deflationary state and prices are falling. Thus, expected return is generally positive. The relationship between the expected return and the real risk-free interest rate, inflation rate, and risk premium can be expressed by the following equation:

$$1 + E(R) = (1 + r_{rF}) \times [1 + E(\pi)] \times [1 + E(RP)]$$

The historical mean return for investment in a particular asset, however, is obtained from the actual return that was earned by an investor. Because the investment is risky, there is no guarantee that the actual return will be equal to the expected return. In fact, it is very unlikely that the two returns are equal for a specific time period being considered. Given a long enough period of time, we can *expect* that the future (expected) return will equal the average historical return. Unfortunately, we do not

5 There are exceptions when an asset reduces overall risk of a portfolio. We will consider those exceptions in Section 14.

know how long that period is—10 years, 50 years, or 100 years. As a practical matter, we often assume that the historical mean return is an adequate representation of the expected return, although this assumption may not be accurate. For example, Exhibit 7 shows that the historical equity returns in the last eight years (2010–2017) for large US company stocks were positive whereas the actual return was negative the prior decade, but nearly always positive historically. Nonetheless, longer-term returns (1926–2017) were positive and could be consistent with expected return. Though it is unknown if the historical mean returns accurately represent expected returns, it is an assumption that is commonly made.

Exhibit 7 Risk and Return for US Asset Classes by Decade (%)

		1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s*	1926–2017
Large company stocks	Return	−0.1	9.2	19.4	7.8	5.9	17.6	18.2	−1.0	13.9	10.2
	Risk	41.6	17.5	14.1	13.1	17.2	19.4	15.9	16.3	13.6	19.8
Small company stocks	Return	1.4	20.7	16.9	15.5	11.5	15.8	15.1	6.3	14.8	12.1
	Risk	78.6	34.5	14.4	21.5	30.8	22.5	20.2	26.1	19.4	31.7
Long-term corporate bonds	Return	6.9	2.7	1	1.7	6.2	13	8.4	7.7	8.3	6.1
	Risk	5.3	1.8	4.4	4.9	8.7	14.1	6.9	11.7	8.8	8.3
Long-term government bonds	Return	4.9	3.2	−0.1	1.4	5.5	12.6	8.8	7.7	6.8	5.5
	Risk	5.3	2.8	4.6	6	8.7	16	8.9	12.4	10.8	9.9
Treasury bills	Return	0.6	0.4	1.9	3.9	6.3	8.9	4.9	2.8	0.2	3.4
	Risk	0.2	0.1	0.2	0.4	0.6	0.9	0.4	0.6	0.1	3.1
Inflation	Return	−2.0	5.4	2.2	2.5	7.4	5.1	2.9	2.5	1.7	2.9
	Risk	2.5	3.1	1.2	0.7	1.2	1.3	0.7	1.6	1.1	4.0

* Through 31 December 2017
Note: Returns are measured as annualized geometric mean returns.
Risk is measured by annualizing monthly standard deviations.
Source: 2018 SBBI Yearbook (Exhibits 1.2,1.3, 2.3 and 6.2).

Going forward, be sure to distinguish between expected return and historical mean return. We will alert the reader whenever historical returns are used to estimate expected returns.

7.2 Nominal Returns of Major US Asset Classes

We focus on three major asset categories in Exhibit 7: stocks, bonds, and T-bills. The mean nominal returns for US asset classes are reported decade by decade since the 1930s. The total for the 1926–2017 period is in the last column. All returns are annual geometric mean returns. Large company stocks had an overall annual return of 10.2 percent during the 92-year period. The return was negative in the 1930s and 2000s, and positive in all remaining decades. The 1950s and 1990s were the best decades for large company stocks. Small company stocks fared even better. The nominal return was never negative for any decade, and had double-digit growth in all decades except two, leading to an overall 92-year annual return of 12.1 percent.

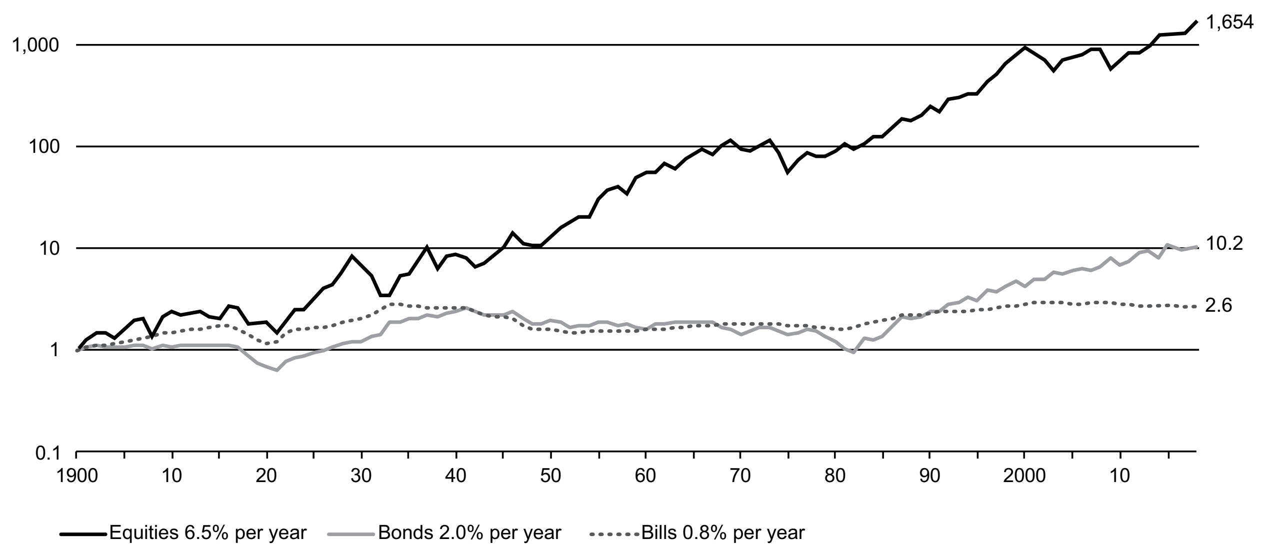
Long-term corporate bonds and long-term government bonds earned overall returns of 6.1 percent and 5.5 percent, respectively. The corporate bonds did not have a single negative decade, although government bonds recorded a negative return in the 1950s when stocks were doing extremely well. Bonds also had some excellent decades, earning double-digit returns in the 1980s and 2000s.

Treasury bills (short-term government securities) did not earn a negative return in any decade. In fact, Treasury bills earned a negative return only in 1938 (–0.02 percent) when the inflation rate was –2.78 percent. Consistently positive returns for Treasury bills are not surprising because nominal interest rates are almost never negative and the Treasury bills suffer from little interest rate or inflation risk. Since the Great Depression, there has been no deflation in any decade, although inflation rates were highly negative in 1930 (–6.03 percent), 1931 (–9.52 percent), and 1932 (–10.30 percent). Conversely, inflation rates were very high in the late 1970s and early 1980s, reaching 13.31 percent in 1979. Inflation rates have been largely range bound between 1 and 3 percent from 1991 to 2017. Overall, the inflation rate was 2.9 percent for the 92-year period.

7.3 Real Returns of Major US Asset Classes

Because annual inflation rates can vary greatly, from –10.30 percent to +13.31 percent in the last 92 years, comparisons across various time periods are difficult and misleading using nominal returns. Therefore, it is more effective to rely on real returns. Real returns on stocks, bonds, and T-bills are reported from 1900 in Exhibits 8 and 9.

Exhibit 8 Cumulative Returns on US Asset Classes in Real Terms, 1900–2017



Source: E. Dimson, P. Marsh, and M. Staunton, *Credit Suisse Global Investment Returns Yearbook 2018*, Credit Suisse Research Institute (February 2018). This chart is updated annually and can be found at https://www.credit-suisse.com/media/assets/corporate/docs/about-us/media/media-release/2018/02/giry-summary-2018.pdf.

Exhibit 8 shows that $1 would have grown to $1,654 if invested in stocks, to only $10.20 if invested in bonds, and to $2.60 if invested in T-bills. The difference in growth among the three asset categories is huge, although the difference in real returns does not seem that large: 6.5 percent per year for equities compared with 2.0 percent per year for bonds. This difference represents the effect of compounding over a 118-year period.

Exhibit 9 reports real rates of return. As we discussed earlier and as shown in the table, geometric mean is never greater than the arithmetic mean. Our analysis of returns focuses on the geometric mean because it is a more accurate representation of returns for multiple holding periods than the arithmetic mean. We observe that the real returns for stocks are higher than the real returns for bonds.

Exhibit 9 Real Returns and Risk Premiums for Asset Classes (1900–2017)

		United States			World			World excluding United States		
	Asset	***GM* (%)**	***AM* (%)**	***SD* (%)**	***GM* (%)**	***AM* (%)**	***SD* (%)**	***GM* (%)**	***AM* (%)**	***SD* (%)**
Real Returns	Equities	6.5	8.4	20.0	5.2	6.6	17.4	4.5	6.2	18.9
	Bonds	2.0	2.5	10.4	2.0	2.5	11.0	1.7	2.7	14.4
Premiums	Equities vs. bonds	4.4	6.5	20.7	3.2	4.4	15.3	2.8	3.8	14.4

Note: All returns are in percent per annum measured in US$. GM = geometric mean, AM = arithmetic mean, SD = standard deviation. "World" consists of 21 developed countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, the Netherlands, New Zealand, Norway, Portugal, South Africa, Spain, Sweden, Switzerland, United Kingdom, and the United States. Weighting is by each country's relative market capitalization size. See source for details of calculations.
Source: Credit Suisse Global Investment Returns Sourcebook, 2018.

7.4 Nominal and Real Returns of Asset Classes in Major Countries

Along with US returns, real returns of major asset classes for a 21-country world and the world excluding the United States are also presented in Exhibit 9. Equity returns are weighted by each country's GDP before 1968 because of a lack of reliable market capitalization data. Returns are weighted by a country's market capitalization beginning with 1968. Similarly, bond returns are defined by a 21-country bond index, except GDP is used to create the weights because equity market capitalization weighting is inappropriate for a bond index and bond market capitalizations were not readily available.

The real geometric mean return for the world stock index over the last 117 years was 5.2 percent, and bonds had a real geometric mean return of 2.0 percent. The real geometric mean return for the world excluding the United States were 4.5 percent for stocks and 1.7 percent for bonds. For both stocks and bonds, the United States earned higher returns than the world excluding the United States. Similarly, real returns for stocks and bonds in the United States were higher than the real returns for rest of the world.

7.5 Risk of Major Asset Classes

Risk for major asset classes in the United States is reported for 1926–2017 in Exhibit 7, and the risk for major asset classes for the United States, the world, and the world excluding the United States are reported for 1900–2017 in Exhibit 9. Exhibit 7 shows that US small company stocks had the highest risk, 31.7 percent, followed by US large company stocks, 19.8 percent. Long-term government bonds and long-term corporate bonds had lower risk at 9.9 percent and 8.3 percent, with Treasury bills having the lowest risk at about 3.1 percent.

Exhibit 9 shows that the risk for world stocks is 17.4 percent and for world bonds is 11.0 percent. The world excluding the United States has risks of 18.9 percent for stocks and 14.4 percent for bonds. The effect of diversification is apparent when world risk is compared with US risk and world excluding US risk. Although the risk of US stocks is 20.0 percent and the risk of world excluding US stocks is 18.9 percent, the combination gives a risk of only 17.4 percent for world stocks.

7.6 Risk–Return Trade-off

The expression "risk–return trade-off" refers to the positive relationship between expected risk and return. In other words, a higher return is not possible to attain in **efficient markets** and over long periods of time without accepting higher risk. Expected returns should be greater for assets with greater risk.

The historical data presented above show the risk–return trade-off. Exhibit 7 shows for the United States that small company stocks had higher risk and higher return than large company stocks. Large company stocks had higher returns and higher risk than both long-term corporate bonds and government bonds. Bonds had higher returns and higher risk than Treasury bills. Uncharacteristically, however, long-term government bonds had higher total risk than long-term corporate bonds, although the returns of corporate bonds were slightly higher. These factors do not mean that long-term government bonds had greater default risk, just that they were more variable than corporate bonds during this historic period.

Exhibit 9 reveals that the risk and return for stocks were the highest of the asset classes, and the risk and return for bonds were lower than stocks for the United States, the world, and the world excluding the United States.

Another way of looking at the risk–return trade-off is to focus on the **risk premium**, which is the extra return investors can expect for assuming additional risk, after accounting for the risk-free interest rate. The nominal risk premium is the nominal risky return minus the nominal risk-free rate (which includes both compensation for expected inflation and the real risk-free interest rate). The real risk premium is the real risky return minus the real risk-free rate. Worldwide equity risk premiums reported at the bottom of Exhibit 9 show that equities outperformed bonds. Investors in equities earned a higher return than investors in bonds because of the higher risk in equities.

A more dramatic representation of the risk–return trade-off is shown in Exhibit 8, which shows the cumulative returns of US asset classes in real terms. The line representing T-bills is much less volatile than the other lines. Adjusted for inflation, the average real return on T-bills was 0.8 percent per year. The line representing bonds is more volatile than the line for T-bills but less volatile than the line representing stocks. The total return for equities including dividends and capital gains shows how $1 invested at the beginning of 1900 grows to $1,654, generating an annualized return of 6.5 percent in real terms.

Over long periods of time, we observe that higher risk does result in higher mean returns. Thus, it is reasonable to claim that, over the long term, market prices reward higher risk with higher returns, which is a characteristic of a risk-averse investor, a topic that we discuss in Section 9.

8 OTHER INVESTMENT CHARACTERISTICS

c describe characteristics of the major asset classes that investors consider in forming portfolios;

In evaluating investments using only the mean (expected return) and variance (risk), we are implicitly making two important assumptions: 1) that the returns are normally distributed and can be fully characterized by their means and variances and 2) that markets are not only informationally efficient but that they are also operationally efficient. To the extent that these assumptions are violated, we need to consider additional investment characteristics. These are discussed below.

8.1 Distributional Characteristics

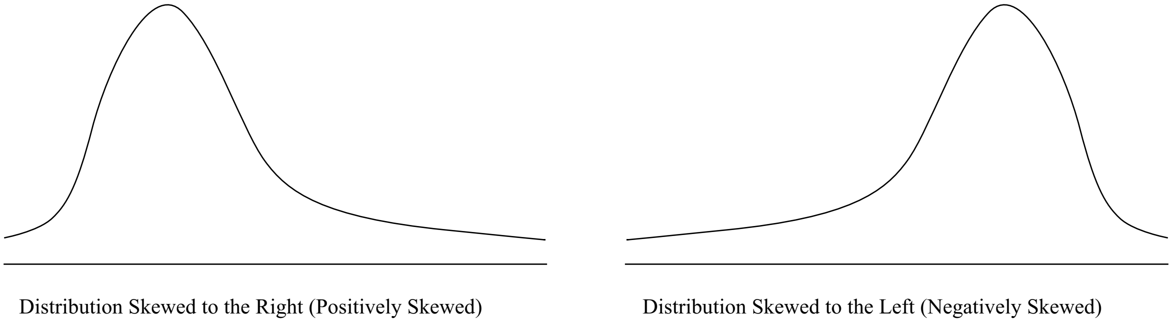
As explained in an earlier reading, a **normal distribution** has three main characteristics: its mean and median are equal; it is completely defined by two parameters, mean and variance; and it is symmetric around its mean with:

- 68 percent of the observations within ±1σ of the mean,
- 95 percent of the observations within ±2σ of the mean, and
- 99 percent of the observations within ±3σ of the mean.

Using only mean and variance would be appropriate to evaluate investments if returns were distributed normally. Returns, however, are not normally distributed; deviations from normality occur both because the returns are skewed, which means they are not symmetric around the mean, and because the probability of extreme events is significantly greater than what a normal distribution would suggest. The latter deviation is referred to as kurtosis or fat tails in a return distribution. The next sections discuss these deviations more in-depth.

Skewness **Skewness** refers to asymmetry of the return distribution, that is, returns are not symmetric around the mean. A distribution is said to be left skewed or negatively skewed if most of the distribution is concentrated to the right, and right skewed or positively skewed if most is concentrated to the left. Exhibit 10 shows a typical representation of negative and positive skewness, whereas Exhibit 11 demonstrates the negative skewness of stock returns by plotting a histogram of US large company stock returns for 1926–2017.

Exhibit 10 Skewness



Source: Reprinted from *Fixed Income Readings for the Chartered Financial Analyst® Program.* Copyright CFA Institute.

Exhibit 11 Histogram of US Large Company Stock Returns, 1926–2017 (Percent)

						2016				
						2014				
						2012				
						2010				
					2015	2006	2017			
					2011	2004	2009	2013		
				2000	2007	1988	2003	1997		
				1990	2005	1986	1999	1995		
				1981	1994	1979	1998	1991		
				1977	1993	1972	1996	1989		
				1969	1992	1971	1983	1985		
				1962	1987	1968	1982	1980		
				1953	1984	1965	1976	1975		
				1946	1978	1964	1967	1955		
			2001	1940	1970	1959	1963	1950		
			1973	1939	1960	1952	1961	1945		
		2002	1966	1934	1956	1949	1951	1938	1958	
	2008	1974	1957	1932	1948	1944	1943	1936	1935	1954
1931	1937	1939	1941	1929	1947	1926	1942	1927	1928	1933

–60 –50 –40 –30 –20 –10 0 10 20 30 40 50 60 70

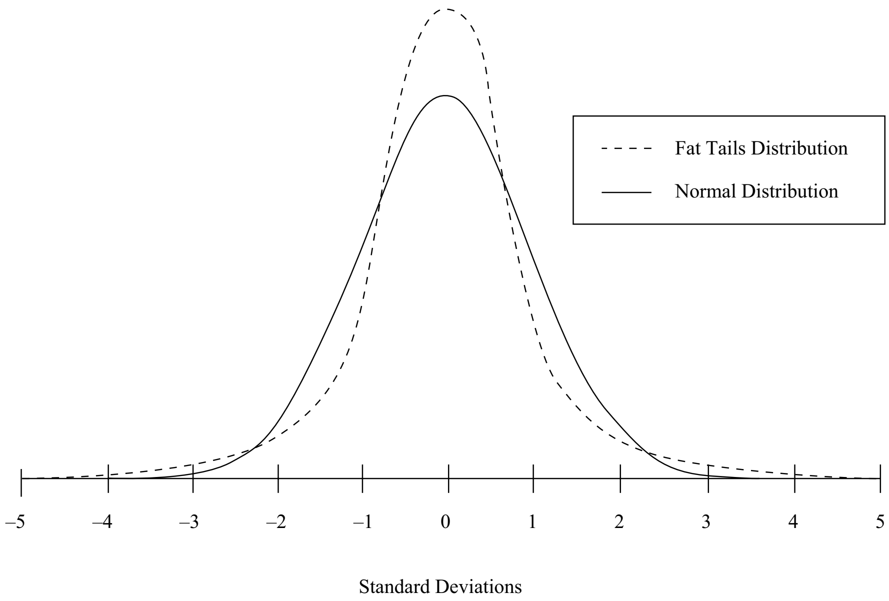
Source: 2018 SBBI Yearbook (Appendix A1)

Kurtosis **Kurtosis** refers to fat tails or higher than normal probabilities for extreme returns and has the effect of increasing an asset's risk that is not captured in a mean–variance framework, as illustrated in Exhibit 12. Investors try to evaluate the effect of kurtosis by using such statistical techniques as value at risk (VaR) and conditional tail expectations.[6] Several market participants note that the probability and the magnitude of extreme events is underappreciated and was a primary contributing factor to the financial crisis of 2008.[7] The higher probability of extreme negative outcomes among stock returns can also be observed in Exhibit 11.

6 Value at risk is a money measure of the minimum losses expected on a portfolio during a specified time period at a given level of probability. It is commonly used to measure the losses a portfolio can suffer under normal market conditions. For example, if a portfolio's one-day 10 percent VaR is £200,000, it implies that there is a 10 percent probability that the value of the portfolio will decrease by more than £200,000 over a single one-day period (under normal market conditions). This probability implies that the portfolio will experience a loss of at least £200,000 on one out of every ten days.

7 For example, see Bogle (2008) and Taleb (2007).

Exhibit 12 Kurtosis



Source: Reprinted from *Fixed Income Readings for the Chartered Financial Analyst® Program.* Copyright CFA Institute.

8.2 Market Characteristics

In the previous analysis, we implicitly assumed that markets are both informationally and operationally efficient. Although informational efficiency of markets is a topic beyond the purview of this reading, we should highlight certain operational limitations of the market that affect the choice of investments. One such limitation is **liquidity**.

The cost of trading has three main components—brokerage commission, bid–ask spread, and price impact. Liquidity affects the latter two. Stocks with low liquidity can have wide bid–ask spreads. The bid–ask spread, which is the difference between the buying price and the selling price, is incurred as a cost of trading a security. The larger the bid–ask spread, the higher the cost of trading. If a $100 stock has a spread of 10 cents, the bid–ask spread is only 0.1 percent ($0.10/$100). On the other hand, if a $10 stock has a spread of 10 cents, the bid–ask spread is 1 percent. Clearly, the $10 stock is more expensive to trade and an investor will need to earn 0.9 percent extra to make up the higher cost of trading relative to the $100 stock.

Liquidity also has implications for the price impact of trade. Price impact refers to how the price moves in response to an order in the market. Small orders usually have little impact, especially for liquid stocks. For example, an order to buy 100 shares of a $100 stock with a spread of 1 cent may have no effect on the price. On the other hand, an order to buy 100,000 shares may have a significant impact on the price as the buyer has to induce more and more stockholders to tender their shares. The extent of the price impact depends on the liquidity of the stock. A stock that trades millions of shares a day may be less affected than a stock that trades only a few hundred thousand shares a day. Investors, especially institutional investors managing large sums of money, must keep the liquidity of a stock in mind when making investment decisions.

Liquidity is a bigger concern in emerging markets than in developed markets because of the smaller volume of trading in those markets. Similarly, liquidity is a more important concern in corporate bond markets and especially for bonds of lower credit quality than in equity markets because an individual corporate bond issue may not trade for several days or weeks. This certainly became apparent during the global financial crisis.

There are other market-related characteristics that affect investment decisions because they might instill greater confidence in the security or might affect the costs of doing business. These include analyst coverage, availability of information, firm size, etc. These characteristics about companies and financial markets are essential components of investment decision making.

RISK AVERSION AND PORTFOLIO SELECTION & THE CONCEPT OF RISK AVERSION

9

e explain risk aversion and its implications for portfolio selection;

As we have seen, stocks, bonds, and T-bills provide different levels of returns and have different levels of risk. Although investment in equities may be appropriate for one investor, another investor may not be inclined to accept the risk that accompanies a share of stock and may prefer to hold more cash. In the last section, we considered investment characteristics of assets in understanding their risk and return. In this section, we consider the characteristics of investors, both individual and institutional, in an attempt to pair the right kind of investors with the right kind of investments.

First, we discuss risk aversion and utility theory. Later we discuss their implications for portfolio selection.

9.1 The Concept of Risk Aversion

The concept of **risk aversion** is related to the behavior of individuals under uncertainty. Assume that an individual is offered two alternatives: one where he will get £50 for sure and the other is a gamble with a 50 percent chance that he gets £100 and 50 percent chance that he gets nothing. The expected value in both cases is £50, one with certainty and the other with uncertainty. What will an investor choose? There are three possibilities: an investor chooses the gamble, the investor chooses £50 with certainty, or the investor is indifferent. Let us consider each in turn. However, please understand that this is only a representative example, and a single choice does not determine the risk aversion of an investor.

Risk Seeking

If an investor chooses the gamble, then the investor is said to be risk loving or risk seeking. The gamble has an uncertain outcome, but with the same expected value as the guaranteed outcome. Thus, an investor choosing the gamble means that the investor gets extra “utility” from the uncertainty associated with the gamble. How much is that extra utility worth? Would the investor be willing to accept a smaller expected value because he gets extra utility from risk? Indeed, risk seekers will accept less return because of the risk that accompanies the gamble. For example, a risk seeker may choose a gamble with an expected value of £45 in preference to a guaranteed outcome of £50.

There is a little bit of gambling instinct in many of us. People buy lottery tickets although the expected value is less than the money they pay to buy it. Or people gamble at casinos with the full knowledge that the expected return is negative, a characteristic of risk seekers. These or any other isolated actions, however, cannot be taken at face value except for compulsive gamblers.

Risk Neutral

If an investor is indifferent about the gamble or the guaranteed outcome, then the investor may be risk neutral. Risk neutrality means that the investor cares only about return and not about risk, so higher return investments are more desirable even if they come with higher risk. Many investors may exhibit characteristics of risk neutrality when the investment at stake is an insignificant part of their wealth. For example, a billionaire may be indifferent about choosing the gamble or a £50 guaranteed outcome.

Risk Averse

If an investor chooses the guaranteed outcome, he/she is said to be **risk averse** because the investor does not want to take the chance of not getting anything at all. Depending on the level of aversion to risk, an investor may be willing to accept a guaranteed outcome of £45 instead of a gamble with an expected value of £50.

In general, investors are likely to shy away from risky investments for a lower, but guaranteed return. That is why they want to minimize their risk for the same amount of return, and maximize their return for the same amount of risk. The risk–return trade-off discussed earlier is an indicator of risk aversion. A risk-neutral investor would maximize return irrespective of risk and a risk-seeking investor would maximize both risk and return.

Data presented in the last section illustrate the historically positive relationship between risk and return, which demonstrates that market prices were based on transactions and investments by risk-averse investors and reflect risk aversion. Therefore, for all practical purposes and for our future discussion, we will assume that the representative investor is a risk-averse investor. This assumption is the standard approach taken in the investment industry globally.

Risk Tolerance

Risk tolerance refers to the amount of risk an investor can tolerate to achieve an investment goal. The higher the risk tolerance, the greater is the willingness to take risk. Thus, risk tolerance is negatively related to risk aversion.

10 UTILITY THEORY AND INDIFFERENCE CURVES

e explain risk aversion and its implications for portfolio selection;

Continuing with our previous example, a risk-averse investor would rank the guaranteed outcome of £50 higher than the uncertain outcome with an expected value of £50. We can say that the utility that an investor or an individual derives from the guaranteed outcome of £50 is greater than the utility or satisfaction or happiness he/she derives from the alternative. In general terms, utility is a measure of relative satisfaction from consumption of various goods and services or in the case of investments, the satisfaction that an investor derives from a portfolio.

Because individuals are different in their preferences, all risk-averse individuals may not rank investment alternatives in the same manner. Consider the £50 gamble again. All risk-averse individuals will rank the guaranteed outcome of £50 higher than the gamble. What if the guaranteed outcome is only £40? Some risk-averse investors might consider £40 inadequate, others might accept it, and still others may now be indifferent about the uncertain £50 and the certain £40.

A simple implementation of utility theory allows us to quantify the rankings of investment choices using risk and return. There are several assumptions about individual behavior that we make in the definition of utility given in the equation below.

We assume that investors are risk averse. They always prefer more to less (greater return to lesser return). They are able to rank different portfolios in the order of their preference and that the rankings are internally consistent. If an individual prefers X to Y and Y to Z, then he/she must prefer X to Z. This property implies that the indifference curves (see Exhibit 13) for the same individual can never touch or intersect. An example of a utility function is given below

$$U = E(r) - \frac{1}{2}A\sigma^2$$

where, U is the utility of an investment, $E(r)$ is the expected return, and σ^2 is the variance of the investment.

In the above equation, A is a measure of risk aversion, which is measured as the marginal reward that an investor requires to accept additional risk. More risk-averse investors require greater compensation for accepting additional risk. Thus, A is higher for more risk-averse individuals. As was mentioned previously, a risk-neutral investor would maximize return irrespective of risk and a risk-seeking investor would maximize both risk and return.

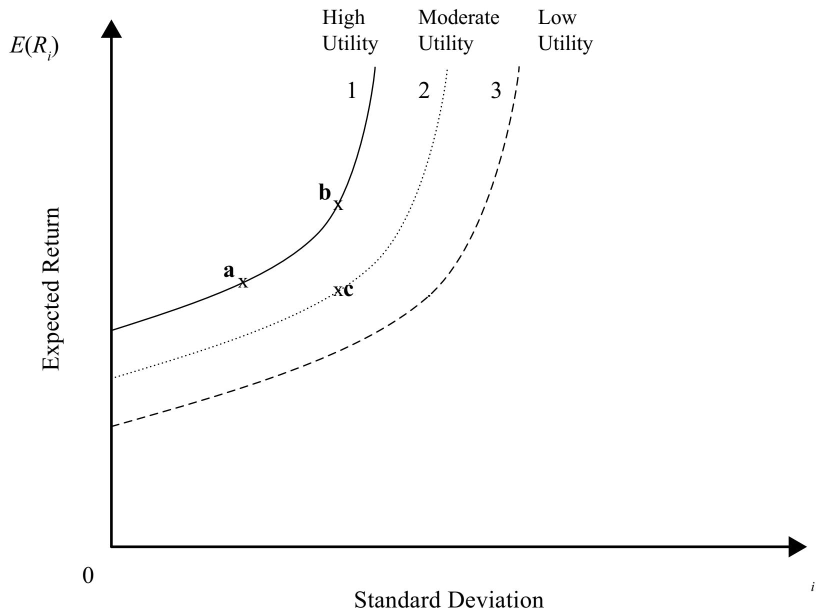
We can draw several conclusions from the utility function. First, utility is unbounded on both sides. It can be highly positive or highly negative. Second, higher return contributes to higher utility. Third, higher variance reduces the utility but the reduction in utility gets amplified by the risk aversion coefficient, A. Utility can always be increased, albeit marginally, by getting higher return or lower risk. Fourth, utility does not indicate or measure satisfaction itself—it can be useful only in ranking various investments. For example, a portfolio with a utility of 4 is not necessarily two times better than a portfolio with a utility of 2. The portfolio with a utility of 4 could increase our happiness 10 times or just marginally. But we do prefer a portfolio with a utility of 4 to a portfolio with a utility of 2. Utility cannot be compared among individuals or investors because it is a very personal concept. From a societal point of view, by the same argument, utility cannot be summed among individuals.

Let us explore the utility function further. The risk aversion coefficient, A, is greater than zero for a risk-averse investor. So any increase in risk reduces his/her utility. The risk aversion coefficient for a risk-neutral investor is 0, and changes in risk do not affect his/her utility. For a risk lover, the risk aversion coefficient is negative, creating an inverse situation so that additional risk contributes to an increase in his/her utility. Note that a risk-free asset ($\sigma^2 = 0$) generates the same utility for all individuals.

10.1 Indifference Curves

An **indifference curve** plots the combinations of risk–return pairs that an investor would accept to maintain a given level of utility (i.e., the investor is indifferent about the combinations on any one curve because they would provide the same level of overall utility). Indifference curves are thus defined in terms of a trade-off between expected rate of return and variance of the rate of return. Because an infinite number of combinations of risk and return can generate the same utility for the same investor, indifference curves are continuous at all points.

Exhibit 13 Indifference Curves for Risk-Averse Investors



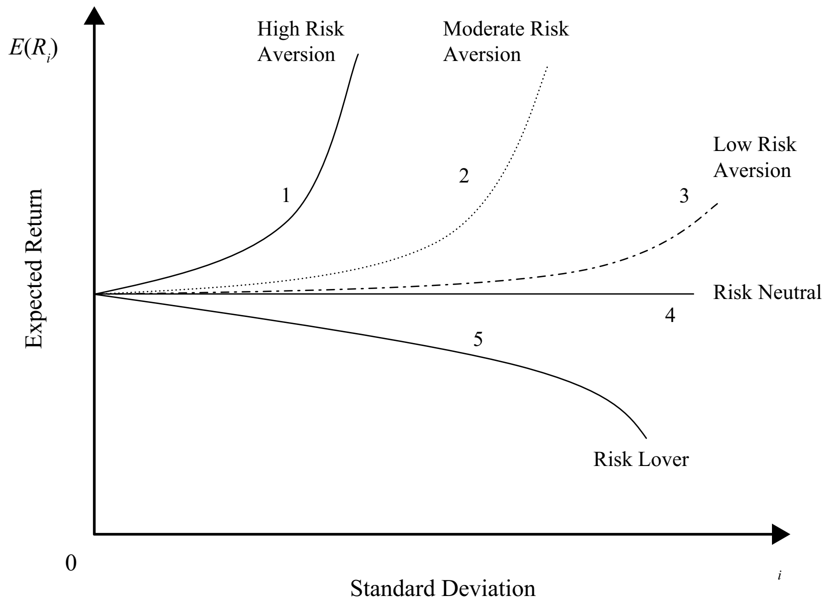
A set of indifference curves is plotted in Exhibit 13. By definition, all points on any one of the three curves have the same utility. An investor does not care whether he/she is at Point **a** or Point **b** on indifference Curve 1. Point **a** has lower risk and lower return than Point **b**, but the utility of both points is the same because the higher return at Point **b** is offset by the higher risk.

Like Curve 1, all points on Curve 2 have the same utility and an investor is indifferent about where he/she is on Curve 2. Now compare Point **c** with Point **b**. Point **c** has the same risk but significantly lower return than Point **b**, which means that the utility at Point **c** is less than the utility at Point **b**. Given that all points on Curve 1 have the same utility and all points on Curve 2 have the same utility and Point **b** has higher utility than Point **c**, Curve 1 has higher utility than Curve 2. Therefore, a risk-averse investor with indifference Curves 1 and 2 will prefer Curve 1 to Curve 2. The utility of a risk-averse investor always increases as you move northwest—higher return with lower risk. Because all investors prefer more utility to less, investors want to move northwest to the indifference curve with the highest utility.

The indifference curve for risk-averse investors runs from the southwest to the northeast because of the risk–return trade-off. If risk increases (going east) then it must be compensated by higher return (going north) to generate the same utility. The indifference curves are convex because of diminishing marginal utility of return (or wealth). As risk increases, an investor needs greater return to compensate for higher risk at an increasing rate (i.e., the curve gets steeper). The upward-sloping convex indifference curve has a slope coefficient closely related to the risk aversion coefficient. The greater the slope, the higher is the risk aversion of the investor as a greater increment in return is required to accept a given increase in risk.

Indifference curves for investors with different levels of risk aversion are plotted in Exhibit 14. The most risk-averse investor has an indifference curve with the greatest slope. As volatility increases, this investor demands increasingly higher returns to compensate for risk. The least risk-averse investor has an indifference curve with the least slope and so the demand for higher return as risk increases is not as acute as for the more risk-averse investor. The risk-loving investor's indifference curve, however, exhibits a negative slope, implying that the risk-lover is happy to substitute risk for return. For a risk lover, the utility increases both with higher risk and higher return. Finally, the indifference curves of risk-neutral investors are horizontal because the utility is invariant with risk.

Exhibit 14 Indifference Curves for Various Types of Investors



In the remaining parts of this reading, all investors are assumed to be risk averse unless stated otherwise.

EXAMPLE 6

Comparing a Gamble with a Guaranteed Outcome

Assume that you are given an investment with an expected return of 10 percent and a risk (standard deviation) of 20 percent, and your risk aversion coefficient is 3.

1. What is your utility of this investment?
2. What must be the minimum risk-free return you should earn to get the same utility?

Solution to 1:

$U = 0.10 - 0.5 \times 3 \times 0.20^2 = 0.04.$

Solution to 2:

A risk-free return's σ is zero, so the second term disappears. To get the same utility (0.04), the risk-free return must be at least 4 percent. Thus, in your mind, a risky return of 10 percent is equivalent to a risk-free return or a guaranteed outcome of 4 percent.

EXAMPLE 7

Computation of Utility

Based on investment information given below and the utility formula $U = E(r) - 0.5A\sigma^2$, answer the following questions. Returns and standard deviations are both expressed as percent per year. When using the utility formula, however, returns and standard deviations must be expressed in decimals.

Investment	Expected Return $E(r)$	Standard Deviation σ
1	12%	30%
2	15	35
3	21	40
4	24	45

1 Which investment will a risk-averse investor with a risk aversion coefficient of 4 choose?

2 Which investment will a risk-averse investor with a risk aversion coefficient of 2 choose?

3 Which investment will a risk-neutral investor choose?

4 Which investment will a risk-loving investor choose?

Solutions to 1 and 2:

The utility for risk-averse investors with A = 4 and A = 2 for each of the four investments are shown in the following table. Complete calculations for Investment 1 with A = 4 are as follows: $U = 0.12 - 0.5 \times 4 \times 0.30^2 = -0.06$.

Investment	Expected Return $E(r)$	Standard Deviation σ	Utility A = 4	Utility A = 2
1	12%	30%	−0.0600	0.0300
2	15	35	−0.0950	0.0275
3	21	40	−0.1100	0.0500
4	24	45	−0.1650	0.0375

The risk-averse investor with a risk aversion coefficient of 4 should choose Investment 1. The risk-averse investor with a risk aversion coefficient of 2 should choose Investment 3.

Solution to 3:

A risk-neutral investor cares only about return. In other words, his risk aversion coefficient is 0. Therefore, a risk-neutral investor will choose Investment 4 because it has the highest return.

Solution to 4:

A risk-loving investor likes both higher risk and higher return. In other words, his risk aversion coefficient is negative. Therefore, a risk-loving investor will choose Investment 4 because it has the highest return and highest risk among the four investments.

APPLICATION OF UTILITY THEORY TO PORTFOLIO SELECTION

11

e explain risk aversion and its implications for portfolio selection;

i explain the selection of an optimal portfolio, given an investor's utility (or risk aversion) and the capital allocation line.

The simplest application of utility theory and risk aversion is to a portfolio of two assets, a risk-free asset and a risky asset. The risk-free asset has zero risk and a return of R_f. The risky asset has a risk of σ_i (> 0) and an expected return of $E(R_i)$. Because the risky asset has risk that is greater than that of the risk-free asset, the expected return from the risky asset will be greater than the return from the risk-free asset, that is, $E(R_i) > R_f$.

We can construct a portfolio of these two assets with a portfolio expected return, $E(R_p)$, and portfolio risk, σ_p, based on the formulas provided below. In the equations given below, w_1 is the weight in the risk-free asset and $(1 - w_1)$ is the weight in the risky asset. Because $\sigma_f = 0$ for the risk-free asset, the first and third terms in the formula for variance are zero leaving only the second term. We arrive at the last equation by taking the square root of both sides, which shows the expression for standard deviation for a portfolio of two assets when one asset is the risk-free asset:

$$E(R_p) = w_1 R_f + (1 - w_1)E(R_i)$$

$$\sigma_P^2 = w_1^2\sigma_f^2 + (1 - w_1)^2\sigma_i^2 + 2w_1(1 - w_1)\rho_{12}\sigma_f\sigma_i = (1 - w_1)^2\sigma_i^2$$

$$\sigma_p = (1 - w_1)\sigma_i$$

The two-asset portfolio is drawn in Exhibit 15 by varying w_1 from 0 percent to 100 percent. The portfolio standard deviation is on the horizontal axis and the portfolio return is on the vertical axis. If only these two assets are available in the economy and the risky asset represents the market, the line in Exhibit 15 is called the **capital allocation line**. The capital allocation line represents the portfolios available to an investor. The equation for this line can be derived from the above two equations by rewriting the second equation as $w_1 = 1 - \dfrac{\sigma_p}{\sigma_i}$. Substituting the value of w_1 in the equation for expected return, we get the following equation for the capital allocation line:

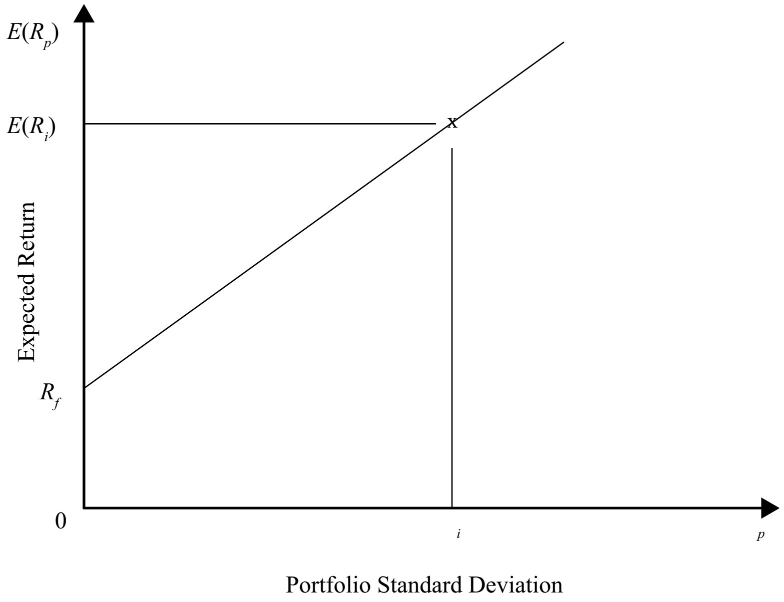
$$E(R_p) = \left(1 - \frac{\sigma_p}{\sigma_i}\right)R_f + \frac{\sigma_p}{\sigma_i}E(R_i)$$

This equation can be rewritten in a more usable form:

$$E(R_p) = R_f + \frac{\left(E(R_i) - R_f\right)}{\sigma_i}\sigma_p$$

The capital allocation line has an intercept of R_f, and a slope of $\dfrac{\left(E(R_i) - R_f\right)}{\sigma_i}$, which is the additional required return for every increment in risk, and is sometimes referred to as the market price of risk.

Exhibit 15 Capital Allocation Line with Two Assets

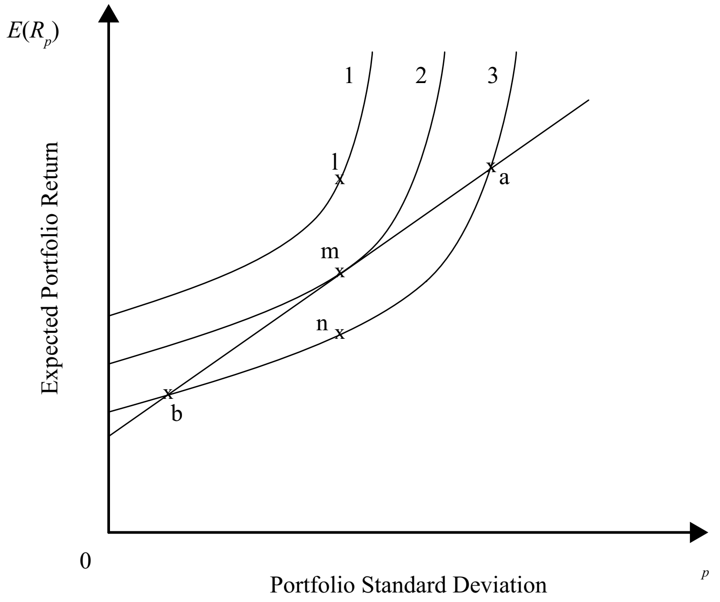


Because the equation is linear, the plot of the capital allocation line is a straight line. The line begins with the risk-free asset as the leftmost point with zero risk and a risk-free return, R_f. At that point, the portfolio consists of only the risk-free asset. If 100 percent is invested in the portfolio of all risky assets, however, we have a return of $E(R_i)$ with a risk of σ_i.

We can move further along the line in pursuit of higher returns by borrowing at the risk-free rate and investing the borrowed money in the portfolio of all risky assets. If 50 percent is borrowed at the risk-free rate, then $w_1 = -0.50$ and 150 percent is placed in the risky asset, giving a return $= 1.50E(R_i) - 0.50R_f$, which is $> E(R_i)$ because $E(R_i) > R_f$.

The line plotted in Exhibit 15 is comprised of an unlimited number of risk–return pairs or portfolios. Which *one* of these portfolios should be chosen by an investor? The answer lies in combining indifference curves from utility theory with the capital allocation line from portfolio theory. Utility theory gives us the utility function or the indifference curves for an individual, as in Exhibit 13, and the capital allocation line gives us the set of feasible investments. Overlaying each individual's indifference curves on the capital allocation line will provide us with the optimal portfolio for that investor. Exhibit 16 illustrates this process of portfolio selection.

Exhibit 16 Portfolio Selection



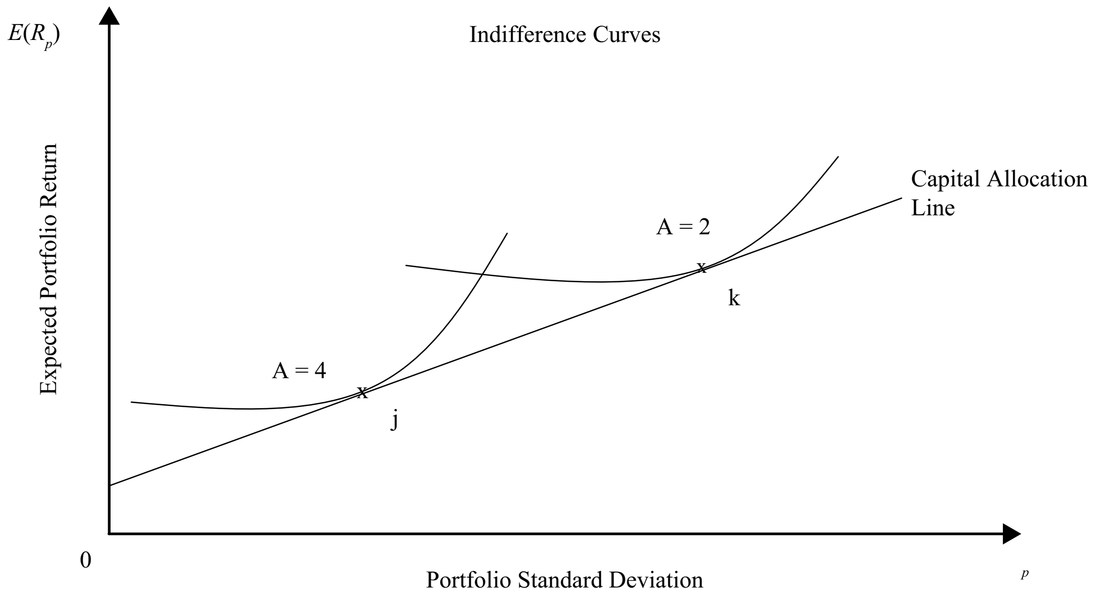
The capital allocation line consists of the set of feasible portfolios. Points under the capital allocation line may be attainable but are not preferred by any investor because the investor can get a higher return for the same risk by moving up to the capital allocation line. Points above the capital allocation line are desirable but not achievable with available assets.

Three indifference curves for the same individual are also shown in Exhibit 16. Curve 1 is above the capital allocation line, Curve 2 is tangential to the line, and Curve 3 intersects the line at two points. Curve 1 has the highest utility and Curve 3 has the lowest utility. Because Curve 1 lies completely above the capital allocation line, points on Curve 1 are not achievable with the available assets on the capital allocation line. Curve 3 intersects the capital allocation line at two Points, **a** and **b**. The investor is able to invest at either Point **a** or **b** to derive the risk–return trade-off and utility associated with Curve 3. Comparing points with the same risk, observe that Point **n** on Curve 3 has the same risk as Point **m** on Curve 2, yet Point **m** has the higher expected return. Therefore, all investors will choose Curve 2 instead of Curve 3. Curve 2 is tangential to the capital allocation line at Point **m**. Point **m** is on the capital allocation line and investable. Point **m** and the utility associated with Curve 2 is the best that the investor can do because he/she cannot move to a higher utility indifference curve. Thus, we have been able to select the optimal portfolio for the investor with indifference Curves 1, 2, and 3. Point **m**, the optimal portfolio for one investor, may not be optimal for another investor. We can follow the same process, however, for finding the optimal portfolio for other investors: the optimal portfolio is the point of tangency between the capital allocation line and the indifference curve for that investor. In other words, the optimal portfolio maximizes the return per unit of risk (as it is on the capital allocation line), and it simultaneously supplies the investor with the most satisfaction (utility).

As an illustration, Exhibit 17 shows two indifference curves for two different investors: Kelly with a risk aversion coefficient of 2 and Jane with a risk aversion coefficient of 4. The indifference curve for Kelly is to the right of the indifference curve for Jane because Kelly is less risk averse than Jane and can accept a higher amount of risk, i.e. has a higher tolerance for risk. Accordingly, their optimal portfolios are different: Point **k** is the optimal portfolio for Kelly and Point **j** is the optimal portfolio for Jane.

In addition, for the same return, the slope of Jane's curve is higher than Kelly's suggesting that Jane needs greater incremental return as compensation for accepting an additional amount of risk compared with Kelly.

Exhibit 17 Portfolio Selection for Two Investors with Various Levels of Risk Aversion



12 PORTFOLIO RISK & PORTFOLIO OF TWO RISKY ASSETS

d calculate and interpret the mean, variance, and covariance (or correlation) of asset returns based on historical data;

f calculate and interpret portfolio standard deviation;

g describe the effect on a portfolio's risk of investing in assets that are less than perfectly correlated;

We have seen before that investors are risk averse and demand a higher return for a riskier investment. Therefore, ways of controlling portfolio risk without affecting return are valuable. As a precursor to managing risk, this section explains and analyzes the components of portfolio risk. In particular, it examines and describes how a portfolio consisting of assets with low correlations have the potential of reducing risk without necessarily reducing return.

12.1 Portfolio of Two Risky Assets

The return and risk of a portfolio of two assets was introduced in Sections 2–8 of this reading. In this section, we briefly review the computation of return and extend the concept of portfolio risk and its components.

12.1.1 *Portfolio Return*

When several individual assets are combined into a portfolio, we can compute the portfolio return as a weighted average of the returns in the portfolio. The portfolio return is simply a weighted average of the returns of the individual investments, or assets. If Asset 1 has a return of 20 percent and constitutes 25 percent of the portfolio's investment, then the contribution to the portfolio return is 5 percent (= 25% of 20%). In general, if Asset i has a return of R_i and has a weight of w_i in the portfolio, then the portfolio return, R_P, is given as:

$$R_P = \sum_{i=1}^{N} w_i R_i, \quad \sum_{i=1}^{N} w_i = 1$$

Note that the weights must add up to 1 because the assets in a portfolio, including cash, must account for 100 percent of the investment. Also, note that these are single period returns, so there are no cash flows during the period and the weights remain constant.

When two individual assets are combined in a portfolio, we can compute the portfolio return as a weighted average of the returns of the two assets. Consider Assets 1 and 2 with weights of 25 percent and 75 percent in a portfolio. If their returns are 20 percent and 5 percent, the weighted average return = (0.25 × 20%) + (0.75 × 5%) = 8.75%. More generally, the portfolio return can be written as below, where R_p is return of the portfolio, w_1 and w_2 are the weights of the two assets, and R_1, R_2 are returns on the two assets:

$$R_p = w_1 R_1 + (1 - w_1) R_2$$

12.1.2 *Portfolio Risk*

Like a portfolio's return, we can calculate a portfolio's variance. Although the return of a portfolio is simply a weighted average of the returns of each security, this is not the case with the standard deviation of a portfolio (unless all securities are perfectly correlated—that is, correlation equals one). Variance can be expressed more generally for N securities in a portfolio using the notation from the portfolio return calculation above:

$$\sum_{i=1}^{N} w_i = 1$$

$$\sigma_P^2 = \text{Var}(R_P) = \text{Var}\left(\sum_{i=1}^{N} w_i R_i\right)$$

Note that the weights must add up to 1. The right side of the equation is the variance of the weighted average returns of individual securities. Weight is a constant, but the returns are variables whose variance is shown by Var(R_i). We can rewrite the equation as shown next. Because the covariance of an asset with itself is the variance of the asset, we can separate the variances from the covariances in the second equation:

$$\sigma_P^2 = \sum_{i,j=1}^{N} w_i w_j \text{Cov}(R_i, R_j)$$

$$\sigma_P^2 = \sum_{i=1}^{N} w_i^2 \text{Var}(R_i) + \sum_{i,j=1, i \neq j}^{N} w_i w_j \text{Cov}(R_i, R_j)$$

Cov(R_i,R_j) is the covariance of returns, R_i and R_j, and can be expressed as the product of the correlation between the two returns ($\rho_{1,2}$) and the standard deviations of the two assets. Thus, Cov(R_i,R_j) = $\rho_{ij}\sigma_i\sigma_j$.

For a two asset portfolio, the expression for portfolio variance simplifies to the following using covariance and then using correlation:

$$\sigma_P^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\text{Cov}(R_1,R_2)$$

$$\sigma_P^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{12}\sigma_1\sigma_2$$

The standard deviation of a two asset portfolio is given by the square root of the portfolio's variance:

$$\sigma_P = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\text{Cov}(R_1,R_2)}$$

or,

$$\sigma_P = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{12}\sigma_1\sigma_2}$$

EXAMPLE 8

Return and Risk of a Two-Asset Portfolio

Assume that as a US investor, you decide to hold a portfolio with 80 percent invested in the S&P 500 US stock index and the remaining 20 percent in the MSCI Emerging Markets index. The expected return is 9.93 percent for the S&P 500 and 18.20 percent for the Emerging Markets index. The risk (standard deviation) is 16.21 percent for the S&P 500 and 33.11 percent for the Emerging Markets index. What will be the portfolio's expected return and risk given that the covariance between the S&P 500 and the Emerging Markets index is 0.5 percent or 0.0050? Note that units for covariance and variance are written as $\%^2$ when not expressed as a fraction. These are units of measure like squared feet and the numbers themselves are not actually squared.

Solution:

Portfolio return, $R_P = w_1R_1 + (1 - w_1), R_2 = (0.80 \times 0.0993) + (0.20 \times 0.1820) = 0.1158 = 11.58\%$.

$$\text{Portfolio risk} = \sigma_P = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2Cov(R_1,R_2)}$$

$$\sigma_p^2 = w_{US}^2\sigma_{US}^2 + w_{EM}^2\sigma_{EM}^2 + 2w_{US}w_{EM}Cov_{US,EM}$$

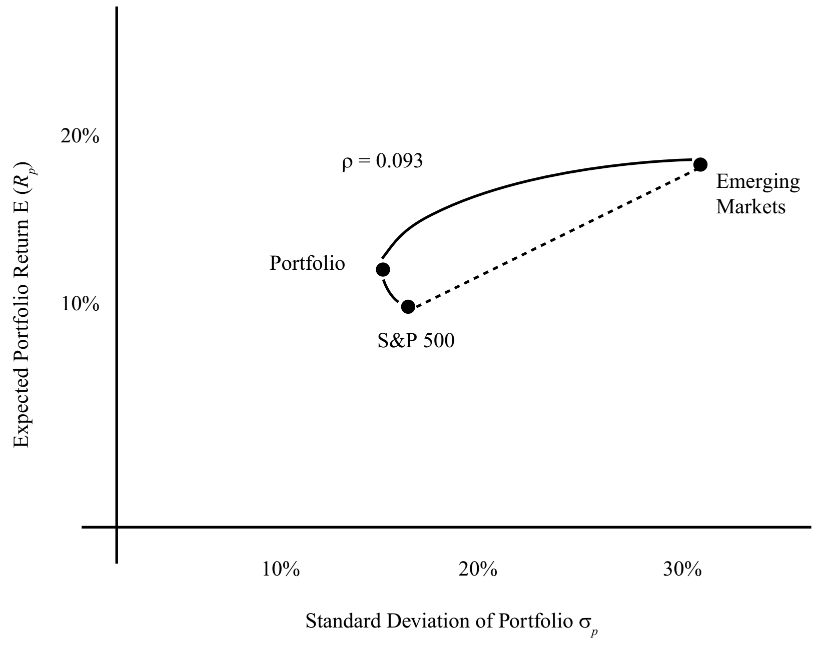
$$\sigma_p^2 = \left(0.80^2 \times 0.1621^2\right) + \left(0.20^2 \times 0.3311^2\right) + \left(2 \times 0.80 \times 0.20 \times 0.0050\right)$$

$$\sigma_p^2 = 0.01682 + 0.00439 + 0.00160 = 0.02281$$

$$\sigma_p = 0.15103 = 15.10\%$$

The portfolio's expected return is 11.58 percent and the portfolio's risk is 15.10 percent. Look at this example closely. It shows that we can take the portfolio of a US investor invested only in the S&P 500, combine it with a *riskier* portfolio consisting of emerging markets securities, and the return of the US investor increases from 9.93 percent to 11.58 percent while the risk of the portfolio actually falls from 16.21 percent to 15.10 percent. Exhibit 18 depicts how the combination of the two assets results in a superior risk–return trade-off. Not only does the investor get a higher return, but he also gets it at a lower risk. That is the power of diversification as you will see later in this reading.

Exhibit 18 Combination of Two Assets



12.1.3 *Covariance and Correlation*

The **covariance** in the formula for portfolio standard deviation can be expanded as $\text{Cov}(R_1,R_2) = \rho_{12}\sigma_1\sigma_2$ where ρ_{12} is the correlation between returns, R_1, R_2. Although covariance is important, it is difficult to interpret because it is unbounded on both sides. It is easier to understand the **correlation coefficient** (ρ_{12}), which is bounded but provides similar information.

Correlation is a measure of the consistency or tendency for two investments to act in a similar way. The correlation coefficient, ρ_{12}, can be positive or negative and ranges from −1 to +1. Consider three different values of the correlation coefficient:

- $\rho_{12} = +1$: Returns of the two assets are perfectly *positively* correlated. Assets 1 and 2 move together 100 percent of the time.
- $\rho_{12} = -1$: Returns of the two assets are perfectly *negatively* correlated. Assets 1 and 2 move in opposite directions 100 percent of the time.
- $\rho_{12} = 0$: Returns of the two assets are *uncorrelated*. Movement of Asset 1 provides no prediction regarding the movement of Asset 2.

The correlation coefficient between two assets determines the effect on portfolio risk when the two assets are combined. To see how this works, consider two different values of ρ_{12}. You will find that portfolio risk is unaffected when the two assets are perfectly correlated ($\rho_{12} = +1$). In other words, the portfolio's standard deviation is simply a weighted average of the standard deviations of the two assets and as such a portfolio's risk is unchanged with the addition of assets with the same risk parameters. Portfolio risk falls, however, when the two assets are not perfectly correlated ($\rho_{12} < +1$). Sufficiently low values of the correlation coefficient can make the portfolio riskless under certain conditions.

First, let ρ_{12} = +1

$$\sigma_p^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{12}\sigma_1\sigma_2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_1\sigma_2$$
$$= \left(w_1\sigma_1 + w_2\sigma_2\right)^2$$
$$\sigma_p = w_1\sigma_1 + w_2\sigma_2$$

The first set of terms on the right side of the first equation contain the usual terms for portfolio variance. Because the correlation coefficient is equal to +1, the right side can be rewritten as a perfect square. The third row shows that portfolio risk is a weighted average of the risks of the individual assets' risks. We showed earlier that the portfolio return is a weighted average of the assets' returns. Because both risk and return are just weighted averages of the two assets in the portfolio there is no reduction in risk when ρ_{12} = +1.

Now let ρ_{12} < +1

The above analysis showed that portfolio risk is a weighted average of asset risks when ρ_{12} = + 1. When ρ_{12} < +1, the portfolio risk is less than the weighted average of the individual assets' risks.

To show this, we begin by reproducing the general formula for portfolio risk, which is expressed by the terms to the left of the "<" sign below. The term to the right of "<" shows the portfolio risk when ρ_{12} = + 1:

$$\sigma_p = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{12}\sigma_1\sigma_2} < \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_1\sigma_2}$$
$$= \left(w_1\sigma_1 + w_2\sigma_2\right)$$
$$\sigma_p < \left(w_1\sigma_1 + w_2\sigma_2\right)$$

The left side is smaller than the right side because the correlation coefficient on the left side for the new portfolio is <1. Thus, the portfolio risk is less than the weighted average of risks while the portfolio return is still a weighted average of returns.

As you can see, we have achieved diversification by combining two assets that are not perfectly correlated. For an extreme case in which ρ_{12} = –1 (that is, the two asset returns move in opposite directions), the portfolio can be made risk free.

EXAMPLE 9

Effect of Correlation on Portfolio Risk

Two stocks have the same return and risk (standard deviation): 10 percent return with 20 percent risk. You form a portfolio with 50 percent each of Stock 1 and Stock 2 to examine the effect of correlation on risk.

1 Calculate the portfolio return and risk if the correlation is 1.0.
2 Calculate the portfolio return and risk if the correlation is 0.0.
3 Calculate the portfolio return and risk if the correlation is –1.0.
4 Compare the return and risk of portfolios with different correlations.

Solution to 1:

$$R_1 = R_2 = 10\% = 0.10;\ \sigma_1 = \sigma_2 = 20\% = 0.20;\ w_1 = w_2 = 50\%$$
$$= 0.50.\ \text{Case 1: } \rho_{12} = +1$$
$$R_p = w_1 R_1 + w_2 R_2$$
$$R_p = (0.5 \times 0.1) + (0.5 \times 0.1) = 0.10 = 10\%$$
$$\sigma_p^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2 w_1 w_2 \sigma_1 \sigma_2 \rho_{12}$$
$$\sigma_p^2 = \left(0.5^2 \times 0.2^2\right) + \left(0.5^2 \times 0.2^2\right) + (2 \times 0.5 \times 0.5 \times 0.2 \times 0.2 \times 1) = 0.04$$
$$\sigma_p = \sqrt{0.04} = 0.20 = 20\%$$

This equation demonstrates the earlier point that with a correlation of 1.0 the risk of the portfolio is the same as the risk of the individual assets.

Solution to 2:

$$\rho_{12} = 0$$
$$R_p = w_1 R_1 + w_2 R_2 = 0.10 = 10\%$$
$$\sigma_p^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2 w_1 w_2 \sigma_1 \sigma_2 \rho_{12}$$
$$\sigma_p^2 = \left(0.5^2 \times 0.2^2\right) + \left(0.5^2 \times 0.2^2\right)$$
$$+ (2 \times 0.5 \times 0.5 \times 0.2 \times 0.2 \times 0) = 0.02$$
$$\sigma_p = \sqrt{0.02} = 0.14 = 14\%$$

This equation demonstrates the earlier point that, when assets have correlations of less than 1.0, they can be combined in a portfolio that has less risk than either of the assets individually.

Solution to 3:

$$\rho_{12} = -1$$
$$R_p = w_1 R_1 + w_2 R_2 = 0.10 = 10\%$$
$$\sigma_p^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2 w_1 w_2 \sigma_1 \sigma_2 \rho_{12}$$
$$\sigma_p^2 = \left(0.5^2 \times 0.2^2\right) + \left(0.5^2 \times 0.2^2\right)$$
$$+ (2 \times 0.5 \times 0.5 \times 0.2 \times 0.2 \times -1) = 0$$
$$\sigma_p = 0\%$$

This equation demonstrates the earlier point that, if the correlation of assets is low enough, in this case 100 percent negative correlation or −1.00 (exactly inversely related), a portfolio can be designed that eliminates risk. The individual assets retain their risk characteristics, but the portfolio is risk free.

Solution to 4:

The expected return is 10 percent in all three cases; however, the returns will be more volatile in Case 1 and least volatile in Case 3. In the first case, there is no diversification of risk (same risk as before of 20 percent) and the return remains the same. In the second case, with a correlation coefficient of 0, we have achieved diversification of risk (risk is now 14 percent instead of 20 percent), again with the same return. In the third case with a correlation coefficient of

–1, the portfolio is risk free, although we continue to get the same return of 10 percent. This example shows the power of diversification that we expand on further in Section 14.

12.1.4 *Relationship between Portfolio Risk and Return*

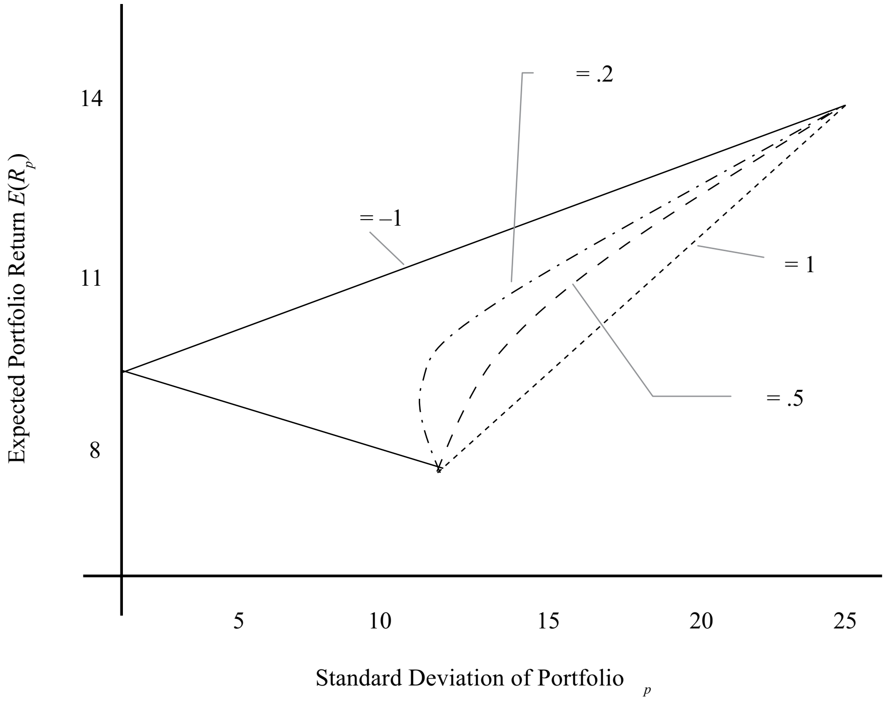
The previous example illustrated the effect of correlation on portfolio risk while keeping the weights in the two assets equal and unchanged. In this section, we consider how portfolio risk and return vary with different portfolio weights and different correlations.

Asset 1 has an annual return of 7 percent and annualized risk of 12 percent, whereas Asset 2 has an annual return of 15 percent and annualized risk of 25 percent. The relationship is tabulated in Exhibit 19 for the two assets and graphically represented in Exhibit 20.

Exhibit 19 Relationship between Risk and Return

Weight in Asset 1 (%)	Portfolio Return	Portfolio Risk with Correlation of 1.0	0.5	0.2	–1.0
0	15.0	25.0	25.0	25.0	25.0
10	14.2	23.7	23.1	22.8	21.3
20	13.4	22.4	21.3	20.6	17.6
30	12.6	21.1	19.6	18.6	13.9
40	11.8	19.8	17.9	16.6	10.2
50	11.0	18.5	16.3	14.9	6.5
60	10.2	17.2	15.0	13.4	2.8
70	9.4	15.9	13.8	12.3	0.9
80	8.6	14.6	12.9	11.7	4.6
90	7.8	13.3	12.2	11.6	8.3
100	7.0	12.0	12.0	12.0	12.0

Exhibit 20 Relationship between Risk and Return



The table shows the portfolio return and risk for four correlation coefficients ranging from +1.0 to –1.0 and 11 weights ranging from 0 percent to 100 percent. The portfolio return and risk are 15 percent and 25 percent, respectively, when 0 percent is invested in Asset 1, versus 7 percent and 12 percent when 100 percent is invested in Asset 1. The portfolio return varies with weights but is unaffected by the correlation coefficient.

Portfolio risk becomes smaller with each successive decrease in the correlation coefficient, with the smallest risk when ρ_{12} = –1. The graph in Exhibit 20 shows that the risk–return relationship is a straight line when ρ_{12} = +1. As the correlation falls, the risk becomes smaller and smaller as in the table. The curvilinear nature of a portfolio of assets is recognizable in all investment opportunity sets (except at the extremes where ρ_{12} = –1 or +1).

EXAMPLE 10

Portfolio of Two Assets

An investor is considering investing in a small-cap stock fund and a general bond fund. Their returns and standard deviations are given below and the correlation between the two fund returns is 0.10.

	Expected Annual Return (%)	Standard Deviation of Returns (%)
Small-cap fund, S	19	33
Bond fund, B	8	13

1 If the investor requires a portfolio return of 12 percent, what should the proportions in each fund be?

2 What is the standard deviation of the portfolio constructed in Part 1?

Solution to 1:

We can calculate the weights by setting the portfolio return equal to 12 percent. 12% = w_1 × 19% + (1 – w_1) × 8%; w_1 = 36.4%, (1 – w_1) = 63.6%. Thus, 36.4 percent should be invested in the small-cap fund and 63.6 percent should be invested in the bond fund.

Solution to 2:

$$\sigma_p = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{12}\sigma_1\sigma_2}$$

$$= \sqrt{\left(0.364^2 \times 0.33^2\right) + \left(0.636^2 \times 0.13^2\right) + (2 \times 0.364 \times 0.636 \times 0.10 \times 0.33 \times 0.13)}$$

$$= 15.23\%$$

The portfolio risk is 15.23 percent, which is much less than a weighted average of risks of 20.28% (63.6% × 13% + 36.4% × 33%).

13 PORTFOLIO OF MANY RISKY ASSETS

f calculate and interpret portfolio standard deviation;

g describe the effect on a portfolio's risk of investing in assets that are less than perfectly correlated;

In the previous section, we discussed how the correlation between two assets can affect the risk of a portfolio and the smaller the correlation the lower is the risk. The above analysis can be extended to a portfolio with many risky assets (N). Recall the previous equations for portfolio return and variance:

$$E\left(R_p\right) = \sum_{i=1}^{N} w_i E\left(R_i\right),\ \ \sigma_P^2 = \left(\sum_{i=1}^{N} w_i^2\sigma_i^2 + \sum_{i,j=1, i\neq j}^{N} w_i w_j \mathrm{Cov}\left(i,j\right)\right),\ \sum_{i=1}^{N} w_i = 1$$

To examine how a portfolio with many risky assets works and the ways in which we can reduce the risk of a portfolio, assume that the portfolio has equal weights (1/N) for all N assets. In addition, assume that $\bar{\sigma}^2$ and $\overline{\mathrm{Cov}}$ are the average variance and average covariance. Given equal weights and average variance/covariance, we can rewrite the portfolio variance as below (intermediate steps are omitted to focus on the main result):

$$\sigma_P^2 = \left(\sum_{i=1}^{N} w_i^2 \sigma_i^2 + \sum_{i,j=1, i \neq j}^{N} w_i w_j \text{Cov}(i,j) \right)$$

$$\sigma_P^2 = \frac{\bar{\sigma}^2}{N} + \frac{(N-1)}{N}\overline{\text{Cov}}$$

The equation in the second line shows that as N becomes large, the first term on the right side with the denominator of N becomes smaller and smaller, implying that the contribution of one asset's variance to portfolio variance gradually becomes negligible. The second term, however, approaches the average covariance as N increases. It is reasonable to say that for portfolios with a large number of assets, covariance among the assets accounts for almost all of the portfolio's risk.

13.1 Importance of Correlation in a Portfolio of Many Assets

The analysis becomes more instructive and interesting if we assume that all assets in the portfolio have the same variance and the same correlation among assets. In that case, the portfolio risk can then be rewritten as:

$$\sigma_p = \sqrt{\frac{\sigma^2}{N} + \frac{(N-1)}{N}\rho\sigma^2}$$

The first term under the root sign becomes negligible as the number of assets in the portfolio increases leaving the second term (correlation) as the main determining factor for portfolio risk. If the assets are unrelated to one another, the portfolio can have close to zero risk. In the next section, we review these concepts to learn how portfolios can be diversified.

THE POWER OF DIVERSIFICATION

14

c describe characteristics of the major asset classes that investors consider in forming portfolios;

g describe the effect on a portfolio's risk of investing in assets that are less than perfectly correlated;

Diversification is one of the most important and powerful concepts in investments. Because investors are risk averse, they are interested in reducing risk preferably without reducing return. In other cases, investors may accept a lower return if it will reduce the chance of catastrophic losses. In previous sections of this reading, you learned the importance of correlation and covariance in managing risk. This section applies those concepts to explore ways for risk diversification. We begin with a simple but intuitive example.

EXAMPLE 11

Diversification with Rain and Shine

Assume a company Beachwear rents beach equipment. The annual return from the company's operations is 20 percent in years with many sunny days but falls to 0 percent in rainy years with few sunny days. The probabilities of a sunny year and a rainy year are equal at 50 percent. Thus, the average return is 10 percent, with a 50 percent chance of 20 percent return and a 50 percent chance of 0 percent return. Because Beachwear can earn a return of 20 percent or 0 percent, its average return of 10 percent is risky.

You are excited about investing in Beachwear but do not like the risk. Having heard about diversification, you decide to add another business to the portfolio to reduce your investment risk.

- There is a snack shop on the beach that sells all the healthy food you like. You estimate that the annual return from the Snackshop is also 20 percent in years with many sunny days and 0 percent in other years. As with the Beachwear shop, the average return is 10 percent.

You decide to invest 50 percent each in Snackshop and Beachwear. The average return is still 10 percent, with 50 percent of 10 percent from Snackshop and 50 percent of 10 percent from Beachwear. In a sunny year, you would earn 20 percent (= 50% of 20% from Beachwear + 50% of 20% from Snackshop). In a rainy year, you would earn 0 percent (=50% of 0% from Beachwear + 50% of 0% from Snackshop). The results are tabulated in Exhibit 21.

Exhibit 21

Type	Company	Percent Invested	Return in Sunny Year (%)	Return in Rainy Year (%)	Average Return (%)
Single stock	Beachwear	100	20	0	10
Single stock	Snackshop	100	20	0	10
Portfolio of two stocks	Beachwear	50	20	0	10
	Snackshop	50	20	0	10
	Total	100	20	0	10

These results seem counterintuitive. You thought that by adding another business you would be able to diversify and reduce your risk, but the risk is exactly the same as before. What went wrong? Note that both businesses do well when it is sunny and both businesses do poorly when it rains. The correlation between the two businesses is +1.0. No reduction in risk occurs when the correlation is +1.0.

- To reduce risk, you must consider a business that does well in a rainy year. You find a company that rents DVDs. DVDrental company is similar to the Beachwear company, except that its annual return is 20 percent in a rainy year and 0 percent in a sunny year, with an average return of 10 percent. DVDrental's 10 percent return is also risky just like Beachwear's return.

If you invest 50 percent each in DVDrental and Beachwear, then the average return is still 10 percent, with 50 percent of 10 percent from DVDrental and 50 percent of 10 percent from Beachwear. In a sunny year, you would earn 10 percent (= 50% of 20% from Beachwear + 50% of 0% from DVDrental). In a rainy year also, you would earn 10 percent (=50% of 0% from Beachwear + 50% of 20% from DVDrental). You have no risk because you earn 10 percent in both sunny and rainy years. Thus, by adding DVDrental to Beachwear, you have reduced (eliminated) your risk without affecting your return. The results are tabulated in Exhibit 22.

Exhibit 22

Type	Company	Percent Invested	Return in Sunny Year (%)	Return in Rainy Year (%)	Average Return (%)
Single stock	Beachwear	100	20	0	10
Single stock	DVDrental	100	0	20	10
Portfolio of two stocks	Beachwear	50	20	0	10
	DVDrental	50	0	20	10
	Total	100	10	10	10

In this case, the two businesses have a correlation of –1.0. When two businesses with a correlation of –1.0 are combined, risk can always be reduced to zero.

14.1 Correlation and Risk Diversification

Correlation is the key in diversification of risk. Notice that the returns from Beachwear and DVDRental always go in the opposite direction. If one of them does well, the other does not. Therefore, adding assets that do not behave like other assets in your portfolio is good and can reduce risk. The two companies in the above example have a correlation of –1.0.

Even when we expand the portfolio to many assets, correlation among assets remains the primary determinant of portfolio risk. Lower correlations are associated with lower risk. Unfortunately, most assets have high positive correlations. The challenge in diversifying risk is to find assets that have a correlation that is much lower than +1.0.

14.2 Historical Risk and Correlation

When we previously discussed asset returns, we were careful to distinguish between historical or past returns and expected or future returns because historical returns may not be a good indicator of future returns. Returns may be highly positive in one period and highly negative in another period depending on the risk of that asset. Exhibit 7 showed that returns for large US company stocks were high in the 1990s but were very low in the 2000s.

Risk for an asset class, however, does not usually change dramatically from one period to the next. Stocks have been risky even in periods of low returns. T-bills are always less risky even when they earn high returns. From Exhibit 7, we can see that risk has typically not varied much from one decade to the next, except that risk for bonds has been much higher in recent decades when compared with earlier decades. Therefore, it is not unreasonable to assume that historical risk can work as a good proxy for future risk.

As with risk, correlations are quite stable among assets of the same country. Intercountry correlations, however, have been on the rise in the last few decades as a result of globalization and the liberalization of many economies. A correlation above 0.90 is considered high because the assets do not provide much opportunity for diversification of risk Low correlations—generally less than 0.50—are desirable for portfolio diversification.

14.3 Historical Correlation among Asset Classes

Correlations among major US asset classes and international stocks are reported in Exhibit 23 for 1970–2017. The highest correlation is between US large company stocks and US small company stocks at about 70 percent, whereas the correlation between US large company stocks and international stocks is approximately 66 percent. Although these are the highest correlations, they still provide diversification benefits because the correlations are less than 100 percent. The correlation between international stocks and US small company stocks is lower, at 50 percent. The lowest correlations are between stocks and bonds, with some correlations being negative, such as that between US small company stocks and US long-term government bonds. Similarly, the correlation between T-bills and stocks is close to zero and is negative for international stocks.[8]

Exhibit 23 Correlation Among US Assets and International Stocks (1970–2017)

Series	International Stocks	US Large Company Stocks	US Small Company Stocks	US Long-Term Corporate Bonds	US Long-Term Treasury Bonds	US T-Bills	US Inflation
International stocks	1.00						
US large company stocks	0.66	1.00					
US small company stocks	0.50	0.72	1.00				
US long-term corporate bonds	0.02	0.23	0.06	1.00			
US long-term Treasury bonds	−0.13	0.01	−0.15	0.89	1.00		
US T-bills	0.01	0.04	0.02	0.05	0.09	1.00	
US inflation	−0.06	−0.11	0.04	−0.32	−0.26	0.69	1.00

Source: 2018 SBBI Yearbook (Exhibit 12.13).

8 In any short period, T-bills are riskless and uncorrelated with other asset classes. For example, a 3-month US Treasury bill is redeemable at its face value upon maturity irrespective of what happens to other assets. When we consider multiple periods, however, returns on T-bills may be related to other asset classes because short-term interest rates vary depending on the strength of the economy and outlook for inflation.

The low correlations between stocks and bonds are attractive for portfolio diversification. Similarly, including international securities in a portfolio can also control portfolio risk. It is not surprising that most diversified portfolios of investors contain domestic stocks, domestic bonds, foreign stocks, foreign bonds, real estate, cash, and other asset classes.

14.4 Avenues for Diversification

The reason for diversification is simple. By constructing a portfolio with assets that do not move together, you create a portfolio that reduces the ups and downs in the short term but continues to grow steadily in the long term. Diversification thus makes a portfolio more resilient to gyrations in financial markets.

We describe a number of approaches for diversification, some of which have been discussed previously and some of which might seem too obvious. Diversification, however, is such an important part of investing that it cannot be emphasized enough, especially when we continue to meet and see many investors who are not properly diversified.

- *Diversify with asset classes.* Correlations among major asset classes[9] are not usually high, as can be observed from the few US asset classes listed in Exhibit 23. Correlations for other asset classes and other countries are also typically low, which provides investors the opportunity to benefit from diversifying among many asset classes to achieve the biggest benefit from diversification. A partial list of asset classes includes domestic large caps, domestic small caps, growth stocks, value stocks, domestic corporate bonds, long-term domestic government bonds, domestic Treasury bills (cash), emerging market stocks, emerging market bonds, developed market stocks (i.e., developed markets excluding domestic market), developed market bonds, real estate, and gold and other commodities. In addition, industries and sectors are used to diversify portfolios. For example, energy stocks may not be well correlated with health care stocks. The exact proportions in which these assets should be included in a portfolio depend on the risk, return, and correlation characteristics of each and the home country of the investor.
- *Diversify with index funds.* Diversifying among asset classes can become costly for small portfolios because of the number of securities required. For example, creating diversified exposure to a single category, such as a domestic large company asset class, may require a group of at least 30 stocks. Exposure to 10 asset classes may require 300 securities, which can be expensive to trade and track. Instead, it may be effective to use exchange-traded funds or mutual funds that track the respective indexes, which could bring down the costs associated with building a well-diversified portfolio. Therefore, many investors should consider index mutual funds as an investment vehicle as opposed to individual securities.
- *Diversification among countries.* Countries are different because of industry focus, economic policy, and political climate. The US economy produces many financial and technical services and invests a significant amount in innovative research. The Chinese and Indian economies, however, are focused on manufacturing. Countries in the European Union are vibrant democracies whereas East Asian countries are experimenting with democracy. Thus, financial returns in one country over time are not likely to be highly correlated with returns in another country. Country returns may also be different because of different currencies. In other words, the return on a foreign investment may be different

9 Major asset classes are distinguished from sub-classes, such as US value stocks and US growth stocks.

when translated to the home country's currency. Because currency returns are uncorrelated with stock returns, they may help reduce the risk of investing in a foreign country even when that country, in isolation, is a very risky emerging market from an equity investment point of view. Investment in foreign countries is an essential part of a well-diversified portfolio.

- *Diversify by* not *owning your employer's stock.* Companies encourage their employees to invest in company stock through employee stock plans and retirement plans. You should evaluate investing in your company, however, just as you would evaluate any other investment. In addition, you should consider the nonfinancial investments that you have made, especially the human capital you have invested in your company. Because you work for your employer, you are already heavily invested in it—your earnings depend on your employer. The level of your earnings, whether your compensation improves or whether you get a promotion, depends on how well your employer performs. If a competitor drives your employer out of the market, you will be out of a job. Additional investments in your employer will concentrate your wealth in one asset even more so and make you less diversified.
- *Evaluate each asset before adding to a portfolio.* Every time you add a security or an asset class to the portfolio, recognize that there is a cost associated with diversification. There is a cost of trading an asset as well as the cost of tracking a larger portfolio. In some cases, the securities or assets may have different names but belong to an asset class in which you already have sufficient exposure. A general rule to evaluate whether a new asset should be included to an existing portfolio is based on the following risk–return trade-off relationship:

$$E(R_{new}) = R_f + \frac{\sigma_{new}\rho_{new,p}}{\sigma_p} \times \left[E(R_p) - R_f\right]$$

  where $E(R)$ is the return from the asset, R_f is the return on the risk-free asset, σ is the standard deviation, ρ is the correlation coefficient, and the subscripts *new* and *p* refer to the new stock and existing portfolio. If the new asset's risk-adjusted return benefits the portfolio, then the asset should be included. The condition can be rewritten using the Sharpe ratio on both sides of the equation as:

$$\frac{E(R_{new}) - R_f}{\sigma_{new}} > \frac{E(R_p) - R_f}{\sigma_p} \times \rho_{new,p}$$

  If the Sharpe ratio of the new asset is greater than the Sharpe ratio of the current portfolio times the correlation coefficient, it is beneficial to add the new asset.
- *Buy insurance for risky portfolios.* It may come as a surprise, but insurance is an investment asset—just a different kind of asset. Insurance has a negative correlation with your assets and is thus very valuable. Insurance gives you a positive return when your assets lose value, but pays nothing if your assets maintain their value. Over time, insurance generates a negative average return. Many individuals, however, are willing to accept a small negative return because insurance reduces their exposure to an extreme loss. In general, it is reasonable to add an investment with a negative return if that investment significantly reduces risk (an example of a classic case of the risk–return trade-off).

Alternatively, investments with negative correlations also exist. Historically, gold has a negative correlation with stocks; however, the expected return is usually small and sometimes even negative. Investors often include gold and other commodities in their portfolios as a way of reducing their overall portfolio risk, including currency risk and inflation risk.

Buying put options is another way of reducing risk. Because put options pay when the underlying asset falls in value (negative correlation), they can protect an investor's portfolio against catastrophic losses. Of course, put options cost money, and the expected return is zero or marginally negative.

EFFICIENT FRONTIER: INVESTMENT OPPORTUNITY SET & MINIMUM VARIANCE PORTFOLIOS 15

g describe the effect on a portfolio's risk of investing in assets that are less than perfectly correlated;

h describe and interpret the minimum-variance and efficient frontiers of risky assets and the global minimum-variance portfolio;

In this section, we formalize the effect of diversification and expand the set of investments to include all available risky assets in a mean–variance framework. The addition of a risk-free asset generates an optimal risky portfolio and the capital allocation line. We can then derive an investor's optimal portfolio by overlaying the capital allocation line with the indifference curves of investors.

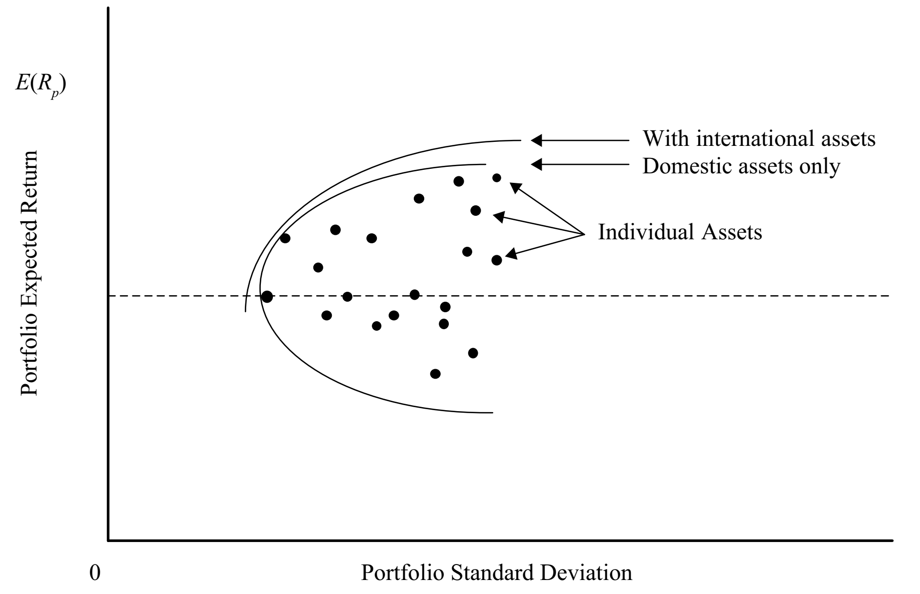
15.1 Investment Opportunity Set

If two assets are perfectly correlated, the risk–return opportunity set is represented by a straight line connecting those two assets. The line contains portfolios formed by changing the weight of each asset invested in the portfolio. This correlation was depicted by the straight line (with $\rho = 1$) in Exhibit 20. If the two assets are not perfectly correlated, the portfolio's risk is less than the weighted average risk of the components, and the portfolio formed from the two assets bulges on the left as shown by curves with the correlation coefficient (ρ) less than 1.0 in Exhibit 20. All of the points connecting the two assets are achievable (or feasible). The addition of new assets to this portfolio creates more and more portfolios that are either a linear combination of the existing portfolio and the new asset or a curvilinear combination, depending on the correlation between the existing portfolio and the new asset.

As the number of available assets increases, the number of possible combinations increases rapidly. When all investable assets are considered, and there are hundreds and thousands of them, we can construct an opportunity set of investments. The opportunity set will ordinarily span all points within a frontier because it is also possible to reach every possible point within that curve by judiciously creating a portfolio from the investable assets.

We begin with individual investable assets and gradually form portfolios that can be plotted to form a curve as shown in Exhibit 24. All points on the curve and points to the right of the curve are attainable by a combination of one or more of the investable assets. This set of points is called the investment opportunity set. Initially, the opportunity set consists of domestic assets only and is labeled as such in Exhibit 24.

Exhibit 24 Investment Opportunity Set



15.1.1 *Addition of Asset Classes*

Exhibit 24 shows the effect of adding a new asset class, such as international assets. As long as the new asset class is not perfectly correlated with the existing asset class, the investment opportunity set will expand out further to the northwest, providing a superior risk–return trade-off.

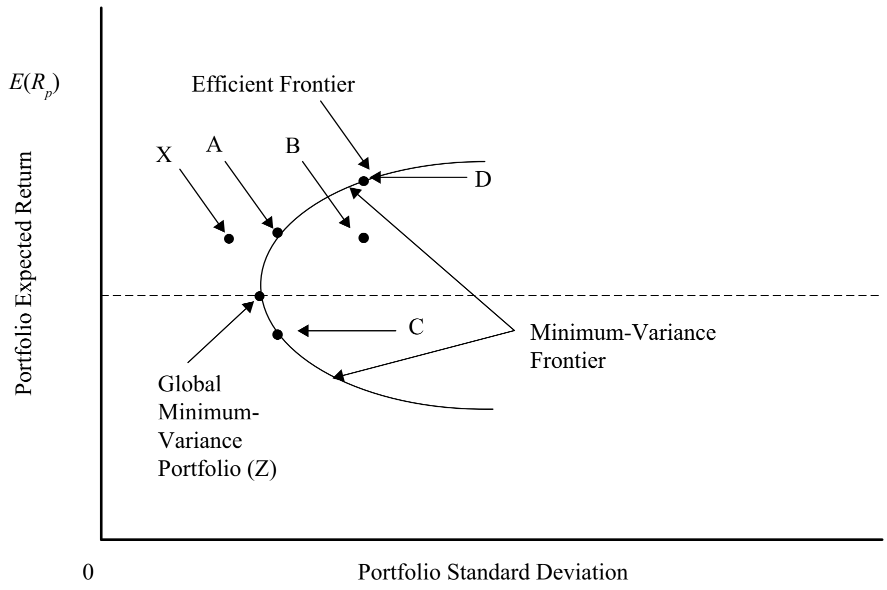
The investment opportunity set with international assets dominates the opportunity set that includes only domestic assets. Adding other asset classes will have the same impact on the opportunity set. Thus, we should continue to add asset classes until they do not further improve the risk–return trade-off. The benefits of diversification can be fully captured in this way in the construction of the investment opportunity set, and eventually in the selection of the optimal portfolio.

In the discussion that follows in this section, we will assume that *all* investable assets available to an investor are included in the investment opportunity set and no special attention needs to be paid to new asset classes or new investment opportunities.

15.2 Minimum-Variance Portfolios

The investment opportunity set consisting of all available investable sets is shown in Exhibit 25. There are a large number of portfolios available for investment, but we must choose a single optimal portfolio. In this subsection, we begin the selection process by narrowing the choice to fewer portfolios.

Exhibit 25 Minimum-Variance Frontier



15.2.1 *Minimum-Variance Frontier*

Risk-averse investors seek to minimize risk for a given return. Consider Points A, B, and X in Exhibit 25 and assume that they are on the same horizontal line by construction. Thus, the three points have the same expected return, $E(R_1)$, as do all other points on the imaginary line connecting A, B, and X. Given a choice, an investor will choose the point with the minimum risk, which is Point X. Point X, however, is unattainable because it does not lie within the investment opportunity set. Thus, the minimum risk that we can attain for $E(R_1)$ is at Point A. Point B and all points to the right of Point A are feasible but they have higher risk. Therefore, a risk-averse investor will choose only Point A in preference to any other portfolio with the same return.

Similarly, Point C is the minimum variance point for the return earned at C. Points to the right of C have higher risk. We can extend the above analysis to all possible returns. In all cases, we find that the **minimum-variance portfolio** is the one that lies on the solid curve drawn in Exhibit 25. The entire collection of these minimum-variance portfolios is referred to as the minimum-variance frontier. The minimum-variance frontier defines the smaller set of portfolios in which investors would want to invest. Note that no risk-averse investor will choose to invest in a portfolio to the right of the minimum-variance frontier because a portfolio on the minimum-variance frontier can give the same return but at a lower risk.

15.2.2 *Global Minimum-Variance Portfolio*

The left-most point on the minimum-variance frontier is the portfolio with the minimum variance among all portfolios of risky assets, and is referred to as the **global minimum-variance portfolio**. An investor cannot hold a portfolio consisting of *risky* assets that has less risk than that of the global minimum-variance portfolio. Note the emphasis on "risky" assets. Later, the introduction of a risk-free asset will allow us to relax this constraint.

15.2.3 *Efficient Frontier of Risky Assets*

The minimum-variance frontier gives us portfolios with the minimum variance for a given return. However, investors also want to maximize return for a given risk. Observe Points A and C on the minimum-variance frontier shown in Exhibit 25. Both of them have the same risk. Given a choice, an investor will choose Portfolio A because it has a higher return. No one will choose Portfolio C. The same analysis applies to all points on the minimum-variance frontier that lie below the global minimum-variance portfolio. Thus, portfolios on the curve below the global minimum-variance portfolio and to the right of the global minimum-variance portfolio are not beneficial and are inefficient portfolios for an investor.

The curve that lies above and to the right of the global minimum-variance portfolio is referred to as the **Markowitz efficient frontier** because it contains all portfolios of risky assets that rational, risk-averse investors will choose.

An important observation that is often ignored is the slope at various points on the efficient frontier. As we move right from the global minimum-variance portfolio (Point Z) in Exhibit 25, there is an increase in risk with a concurrent increase in return. The increase in return with every unit increase in risk, however, keeps decreasing as we move from left to the right because the slope continues to decrease. The slope at Point D is less than the slope at Point A, which is less than the slope at Point Z. The increase in return by moving from Point Z to Point A is the same as the increase in return by moving from Point A to Point D. It can be seen that the additional risk in moving from Point A to Point D is 3 to 4 times more than the additional risk in moving from Point Z to Point A. Thus, investors obtain decreasing increases in returns as they assume more risk.

16 EFFICIENT FRONTIER: A RISK-FREE ASSET AND MANY RISKY ASSETS

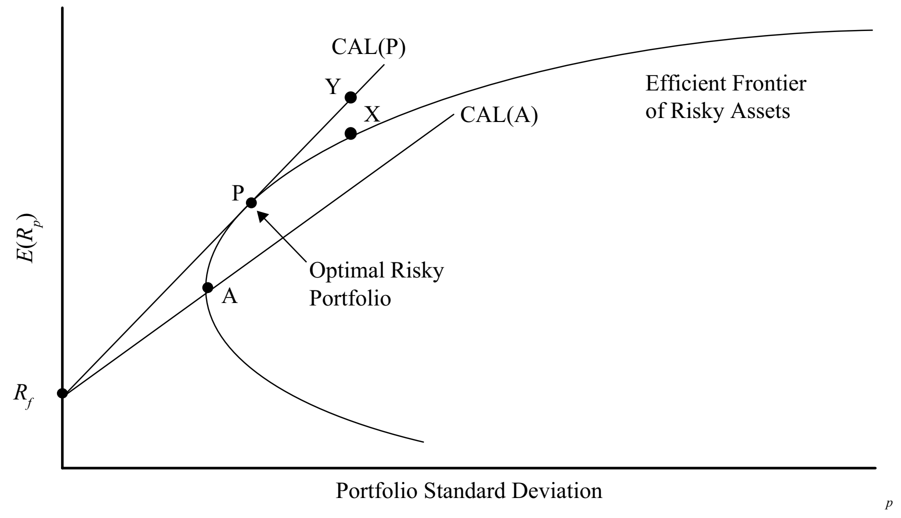
i. explain the selection of an optimal portfolio, given an investor's utility (or risk aversion) and the capital allocation line.

Until now, we have only considered risky assets in which the return is risky or uncertain. Most investors, however, have access to a risk-free asset, most notably from securities issued by the government. The addition of a risk-free asset makes the investment opportunity set much richer than the investment opportunity set consisting only of risky assets.

16.1 Capital Allocation Line and Optimal Risky Portfolio

By definition, a risk-free asset has zero risk so it must lie on the y-axis in a mean-variance graph. A risk-free asset with a return of R_f is plotted in Exhibit 26. This asset can now be combined with a portfolio of risky assets. The combination of a risk-free asset with a portfolio of risky assets is a straight line, such as in Section 11 (see Exhibit 15). Unlike in Section 11, however, we have many risky portfolios to choose from instead of a single risky portfolio.

Exhibit 26 Optimal Risky Portfolio



All portfolios on the efficient frontier are candidates for being combined with the risk-free asset. Two combinations are shown in Exhibit 26: one between the risk-free asset and efficient Portfolio A and the other between the risk-free asset and efficient Portfolio P. Comparing capital allocation line A and capital allocation line P reveals that there is a point on CAL(P) with a higher return and same risk for each point on CAL(A). In other words, the portfolios on CAL(P) dominate the portfolios on CAL(A). Therefore, an investor will choose CAL(P) over CAL(A). We would like to move further northwest to achieve even better portfolios. None of those portfolios, however, is attainable because they are above the efficient frontier.

What about other points on the efficient frontier? For example, Point X is on the efficient frontier and has the highest return of all risky portfolios for its risk. However, Point Y on CAL(P), achievable by leveraging Portfolio P as seen in Section 11, lies above Point X and has the same risk but higher return. In the same way, we can observe that not only does CAL(P) dominate CAL(A) but it also dominates the Markowitz efficient frontier of risky assets.

CAL(P) is the optimal capital allocation line and Portfolio P is the optimal risky portfolio. Thus, with the addition of the risk-free asset, we are able to narrow our selection of risky portfolios to a single optimal risky portfolio, P, which is at the tangent of CAL(P) and the efficient frontier of risky assets.

16.2 The Two-Fund Separation Theorem

The **two-fund separation theorem** states that all investors regardless of taste, risk preferences, and initial wealth will hold a combination of two portfolios or funds: a risk-free asset and an optimal portfolio of risky assets.[10]

The separation theorem allows us to divide an investor's investment problem into two distinct steps: the investment decision and the financing decision. In the first step, as in the previous analysis, the investor identifies the optimal risky portfolio. The optimal risky portfolio is selected from numerous risky portfolios without considering the investor's preferences. The investment decision at this step is based on the optimal risky portfolio's (a single portfolio) return, risk, and correlations.

10 In the next reading, you will learn that the optimal portfolio of risky assets is the market portfolio.

The capital allocation line connects the optimal risky portfolio and the risk-free asset. All optimal investor portfolios must be on this line. Each investor's optimal portfolio on the CAL(P) is determined in the second step. Considering each individual investor's risk preference, using indifference curves, determines the investor's allocation to the risk-free asset (lending) and to the optimal risky portfolio. Portfolios beyond the optimal risky portfolio are obtained by borrowing at the risk-free rate (i.e., buying on margin). Therefore, the individual investor's risk preference determines the amount of financing (i.e., lending to the government instead of investing in the optimal risky portfolio or borrowing to purchase additional amounts of the optimal risky portfolio).

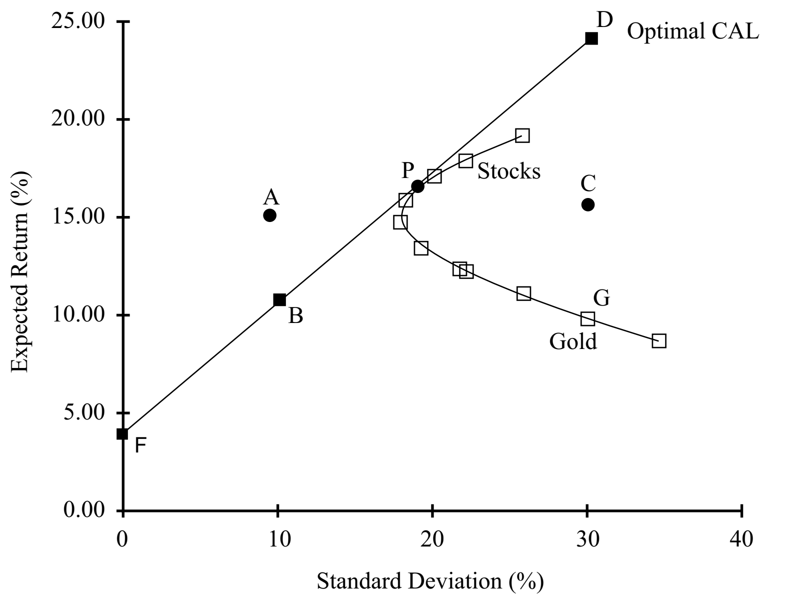
EXAMPLE 12

Choosing the Right Portfolio

In Exhibit 27, the risk and return of the points marked are as follows:

Point	Return (%)	Risk (%)	Point (%)	Return (%)	Risk (%)
A	15	10	B	11	10
C	15	30	D	25	30
F	4	0	G (gold)	10	30
P	16	17			

Exhibit 27



Answer the following questions with reference to the points plotted on Exhibit 27 and explain your answers. The investor is choosing one portfolio based on the graph.

1 Which of the above points is not achievable?

2 Which of these portfolios will not be chosen by a rational, risk-averse investor?

3 Which of these portfolios is most suitable for a risk-neutral investor?

4 Gold is on the inefficient part of the feasible set. Nonetheless, gold is owned by many rational investors as part of a larger portfolio. Why?

5 What is the utility of an investor at point P with a risk aversion coefficient of 3?

Solution to 1:

Portfolio A is not attainable because it lies outside the feasible set and not on the capital allocation line.

Solution to 2:

Portfolios G and C will not be chosen because D provides higher return for the same risk. G and C are the only investable points that do not lie on the capital allocation line.

Solution to 3:

Portfolio D is most suitable because a risk-neutral investor cares only about return and portfolio D provides the highest return. A = 0 in the utility formula.

Solution to 4:

Gold may be owned as part of a portfolio (not as *the* portfolio) because gold has low or negative correlation with many risky assets, such as stocks. Being part of a portfolio can thus reduce overall risk even though its standalone risk is high and return is low. Note that gold's price is not stable—its return is very risky (30 percent). Even risk seekers will choose D over G, which has the same risk but higher return.

Solution to 5:

$$U = E(r) - 0.5A\sigma^2 = 0.16 - 0.5 \times 3 \times 0.0289 = 0.1167 = 11.67\%.$$

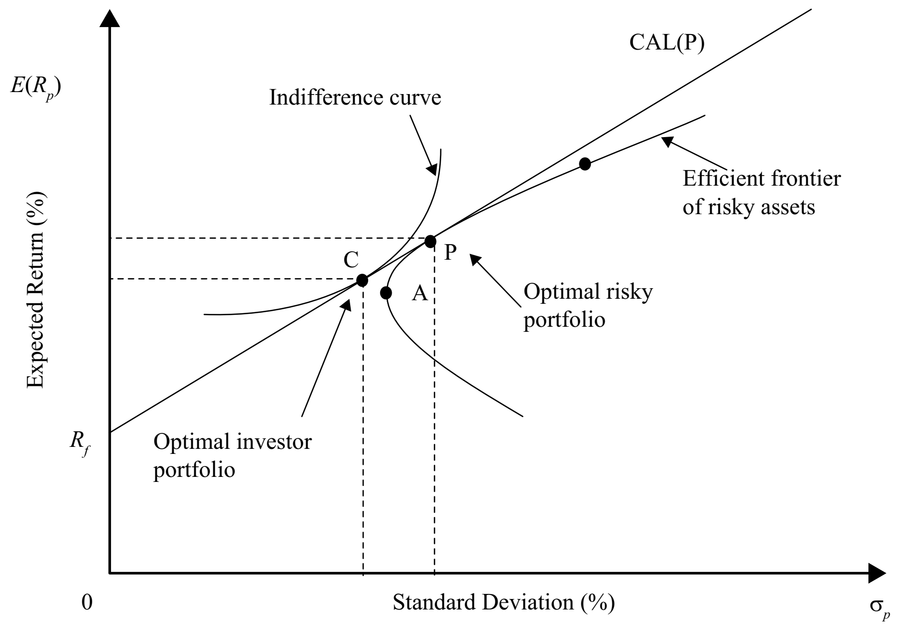
EFFICIENT FRONTIER: OPTIMAL INVESTOR PORTFOLIO 17

i. explain the selection of an optimal portfolio, given an investor's utility (or risk aversion) and the capital allocation line.

The CAL(P) in Exhibits 26 and 28 contains the best possible portfolios available to investors. Each of those portfolios is a linear combination of the risk-free asset and the optimal risky portfolio. Among the available portfolios, the selection of each investor's optimal portfolio depends on the risk preferences of an investor. In Sections 9–11, we discussed that the individual investor's risk preferences are incorporated into their indifference curves. These can be used to select the optimal portfolio.

Exhibit 28 shows an indifference curve that is tangent to the capital allocation line, CAL(P). Indifference curves with higher utility than this one lie above the capital allocation line, so their portfolios are not achievable. Indifference curves that lie below this one are not preferred because they have lower utility. Thus, the optimal portfolio for the investor with this indifference curve is portfolio C on CAL(P), which is tangent to the indifference curve.

Exhibit 28 Optimal Investor Portfolio



EXAMPLE 13

Comprehensive Example on Portfolio Selection

This comprehensive example reviews many concepts learned in this reading. The example begins with simple information about available assets and builds an optimal investor portfolio for the Lohrmanns.

Suppose the Lohrmanns can invest in only two risky assets, A and B. The expected return and standard deviation for asset A are 20 percent and 50 percent, and the expected return and standard deviation for asset B are 15 percent and 33 percent. The two assets have zero correlation with one another.

1 Calculate portfolio expected return and portfolio risk (standard deviation) if an investor invests 10 percent in A and the remaining 90 percent in B.

Solution to 1:

The subscript "*rp*" means risky portfolio.

$$R_{rp} = [0.10 \times 20\%] + [(1 - 0.10) \times 15\%] = 0.155 = 15.50\%$$

$$\begin{aligned}\sigma_{rp} &= \sqrt{w_A^2\sigma_A^2 + w_B^2\sigma_B^2 + 2w_A w_B \rho_{AB}\sigma_A\sigma_B} \\ &= \sqrt{\left(0.10^2 \times 0.50^2\right) + \left(0.90^2 \times 0.33^2\right) + (2 \times 0.10 \times 0.90 \times 0.0 \times 0.50 \times 0.33)} \\ &= 0.3012 = 30.12\%\end{aligned}$$

Note that the correlation coefficient is 0, so the last term for standard deviation is zero.

2 Generalize the above calculations for portfolio return and risk by assuming an investment of w_A in Asset A and an investment of $(1 - w_A)$ in Asset B.

Solution to 2:

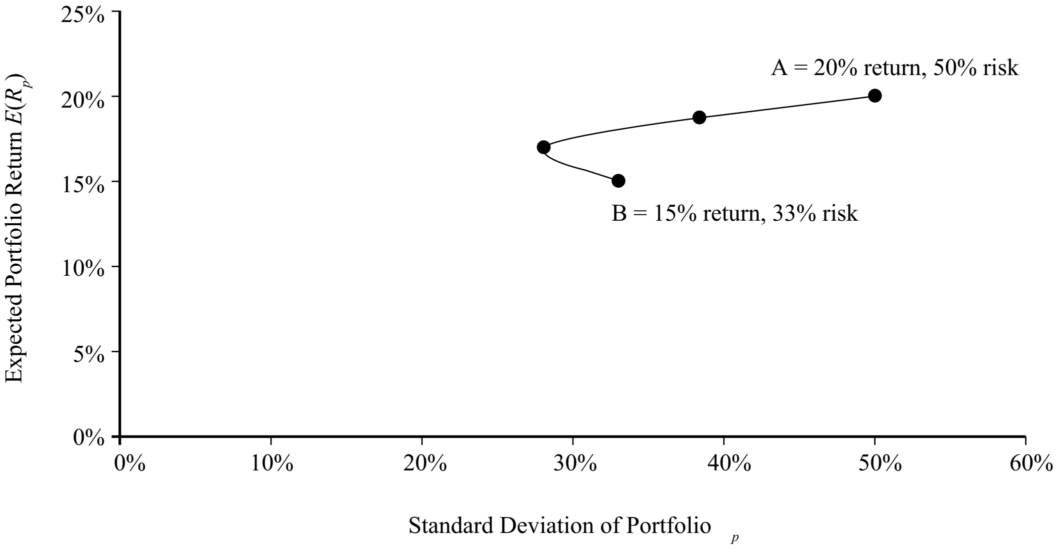
$$R_{rp} = w_A \times 20\% + (1 - w_A) \times 15\% = 0.05w_A + 0.15$$

$$\sigma_{rp} = \sqrt{w_A^2 \times 0.5^2 + (1 - w_A)^2 \times 0.33^2} = \sqrt{0.25w_A^2 + 0.1089(1 - 2w_A + w_A^2)}$$

$$= \sqrt{0.3589w_A^2 - 0.2178w_A + 0.1089}$$

The investment opportunity set can be constructed by using different weights in the expressions for $E(R_{rp})$ and σ_{rp} in Part 1 of this example. Exhibit 29 shows the combination of Assets A and B.

Exhibit 29



3 Now introduce a risk-free asset with a return of 3 percent. Write an equation for the capital allocation line in terms of w_A that will connect the risk-free asset to the portfolio of risky assets. (Hint: use the equation in Section 11 and substitute the expressions for a risky portfolio's risk and return from Part 2 above).

Solution to 3:

The equation of the line connecting the risk-free asset to the portfolio of risky assets is given below (see Section 11), where the subscript "*rp*" refers to the risky portfolio instead of "*i*," and the subscript "*p*" refers to the new portfolio of two risky assets and one risk-free asset.

$$E(R_p) = R_f + \frac{E(R_i) - R_f}{\sigma_i}\sigma_p,$$

Rewritten as

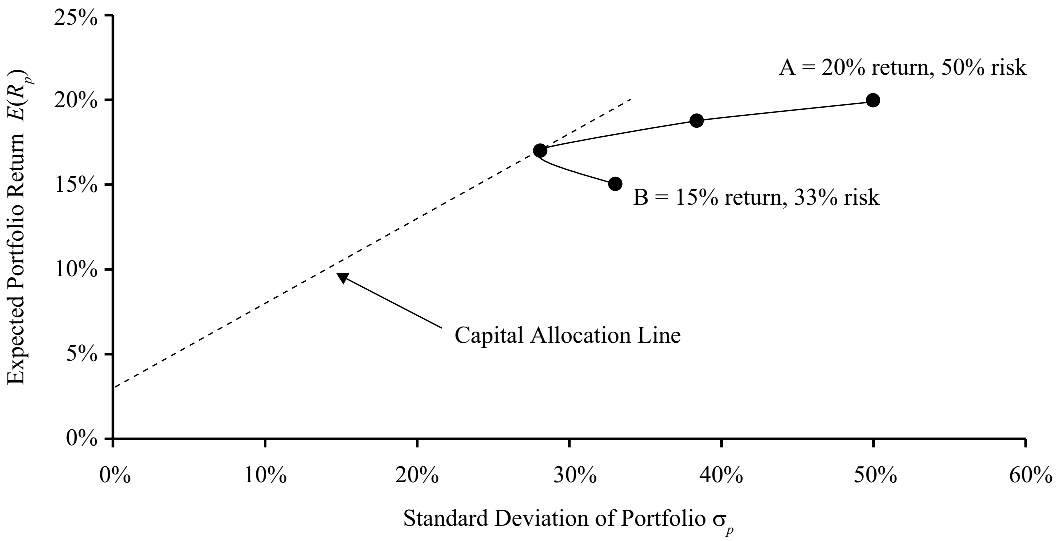
$$E(R_p) = R_f + \frac{E(R_{rp}) - R_f}{\sigma_{rp}}\sigma_p$$

$$= 0.03 + \frac{0.05w_A + 0.15 - 0.03}{\sqrt{0.3589w_A^2 - 0.2178w_A + 0.1089}}\sigma_p$$

$$= 0.03 + \frac{0.05w_A + 0.12}{\sqrt{0.3589w_A^2 - 0.2178w_A + 0.1089}}\sigma_p$$

The capital allocation line is the line that has the maximum slope because it is tangent to the curve formed by portfolios of the two risky assets. Exhibit 30 shows the capital allocation line based on a risk-free asset added to the group of assets.

Exhibit 30



4 The slope of the capital allocation line is maximized when the weight in Asset A is 38.20 percent.[11] What is the equation for the capital allocation line using w_A of 38.20 percent?

11 You can maximize $\frac{0.05w_A + 0.12}{\sqrt{0.3589w_A^2 - 0.2178w_A + 0.1089}}$ by taking the first derivative of the slope with respect to w_A and setting it to 0.

Solution to 4:

By substituting 38.20 percent for w_A in the equation in Part 3, we get $E(R_p) = 0.03 + 0.4978\sigma_p$ as the capital allocation line.

5 Having created the capital allocation line, we turn to the Lohrmanns. What is the standard deviation of a portfolio that gives a 20 percent return and is on the capital allocation line? How does this portfolio compare with asset A?

Solution to 5:

Solve the equation for the capital allocation line to get the standard deviation: $0.20 = 0.03 + 0.4978\sigma_p$. $\sigma_p = 34.2\%$. The portfolio with a 20 percent return has the same return as Asset A but a lower standard deviation, 34.2 percent instead of 50.0 percent.

6 What is the risk of portfolios with returns of 3 percent, 9 percent, 15 percent, and 20 percent?

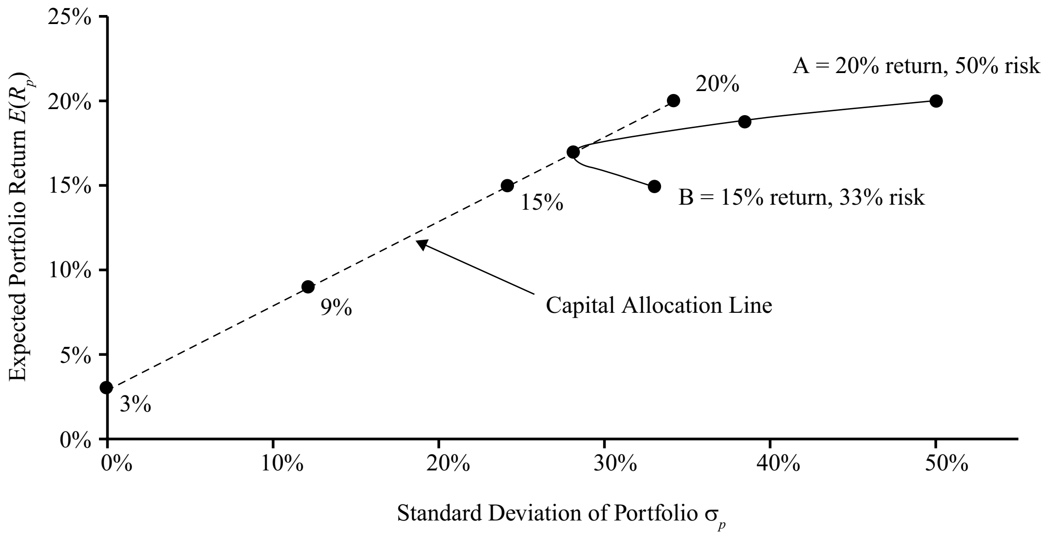
Solution to 6:

You can find the risk of the portfolio using the equation for the capital allocation line: $E(R_p) = 0.03 + 0.4978\sigma_p$.

For a portfolio with a return of 15 percent, write $0.15 = 0.03 + 0.4978\sigma_p$. Solving for σ_p gives 24.1 percent. You can similarly calculate risks of other portfolios with the given returns.

The risk of the portfolio for a return of 3 percent is 0.0 percent, for a return of 9 percent is 12.1 percent, for a return of 15 percent is 24.1 percent, and for a return of 20 percent is 34.2 percent. The points are plotted in Exhibit 31.

Exhibit 31



7 What is the utility that the Lohrmanns derive from a portfolio with a return of 3 percent, 9 percent, 15 percent, and 20 percent? The risk aversion coefficient for the Lohrmanns is 2.5.

Solution to 7:

To find the utility, use the utility formula with a risk aversion coefficient of 2.5:

$$\text{Utility} = E\left(R_p\right) - 0.5 \times 2.5\sigma_p^2$$

$$\text{Utility}\left(3\%\right) = 0.0300$$

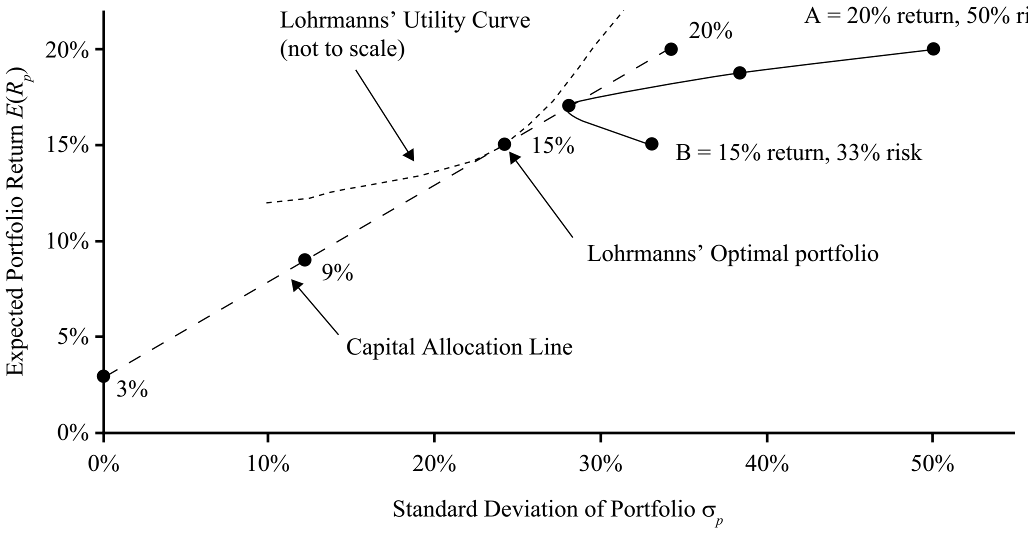
$$\text{Utility}\left(9\%\right) = 0.09 - 0.5 \times 2.5 \times 0.121^2 = +0.0717$$

$$\text{Utility}\left(15\%\right) = 0.15 - 0.5 \times 2.5 \times 0.241^2 = +0.0774$$

$$\text{Utility}\left(20\%\right) = 0.20 - 0.5 \times 2.5 \times 0.341^2 = +0.0546$$

Based on the above information, the Lohrmanns choose a portfolio with a return of 15 percent and a standard deviation of 24.1 percent because it has the highest utility: 0.0774. Finally, Exhibit 32 shows the indifference curve that is tangent to the capital allocation line to generate Lohrmanns' optimal investor portfolio.

Exhibit 32



17.1 Investor Preferences and Optimal Portfolios

The location of an optimal investor portfolio depends on the investor's risk preferences. A highly risk-averse investor may invest a large proportion, even 100 percent, of his/her assets in the risk-free asset. The optimal portfolio in this investor's case will be located close to the y-axis. A less risk-averse investor, however, may invest a large portion of his/her wealth in the optimal risky asset. The optimal portfolio in this investor's case will lie closer to Point P in Exhibit 28.

Some less risk-averse investors (i.e., with a high risk tolerance) may wish to accept even more risk because of the chance of higher return. Such an investor may borrow money to invest more in the risky portfolio. If the investor borrows 25 percent of his wealth, he/she can invest 125 percent in the optimal risky portfolio. The optimal investor portfolio for such an investor will lie to the right of Point P on the capital allocation line.

Thus, moving from the risk-free asset along the capital allocation line, we encounter investors who are willing to accept more risk. At Point P, the investor is 100 percent invested in the optimal risky portfolio. Beyond Point P, the investor accepts even more risk by borrowing money and investing in the optimal risky portfolio.

Note that we are able to accommodate all types of investors with just two portfolios: the risk-free asset and the optimal risky portfolio. Exhibit 28 is also an illustration of the two-fund separation theorem. Portfolio P is the optimal risky portfolio that is selected without regard to investor preferences. The optimal investor portfolio is selected on the capital allocation line by overlaying the indifference curves that incorporate investor preferences.

SUMMARY

This reading provides a description and computation of investment characteristics, such as risk and return, that investors use in evaluating assets for investment. This was followed by sections about portfolio construction, selection of an optimal risky portfolio, and an understanding of risk aversion and indifference curves. Finally, the tangency point of the indifference curves with the capital allocation line allows identification of the optimal investor portfolio. Key concepts covered in the reading include the following:

- Holding period return is most appropriate for a single, predefined holding period.
- Multiperiod returns can be aggregated in many ways. Each return computation has special applications for evaluating investments.
- Risk-averse investors make investment decisions based on the risk–return trade-off, maximizing return for the same risk, and minimizing risk for the same return. They may be concerned, however, by deviations from a normal return distribution and from assumptions of financial markets' operational efficiency.
- Investors are risk averse, and historical data confirm that financial markets price assets for risk-averse investors.
- The risk of a two-asset portfolio is dependent on the proportions of each asset, their standard deviations and the correlation (or covariance) between the assets' returns. As the number of assets in a portfolio increases, the correlation among asset risks becomes a more important determinate of portfolio risk.
- Combining assets with low correlations reduces portfolio risk.
- The two-fund separation theorem allows us to separate decision making into two steps. In the first step, the optimal risky portfolio and the capital allocation line are identified, which are the same for all investors. In the second step, investor risk preferences enable us to find a unique optimal investor portfolio for each investor.
- The addition of a risk-free asset creates portfolios that are dominant to portfolios of risky assets in all cases except for the optimal risky portfolio.

By successfully understanding the content of this reading, you should be comfortable calculating an investor's optimal portfolio given the investor's risk preferences and universe of investable assets available.

REFERENCES

2009 Ibbotson Stocks, Bonds, Bills, and Inflation (SBBI) Classic Yearbook. 2009. Chicago, IL: Morningstar.

Bogle, John C. 2008. “Black Monday and Black Swans.” *Financial Analysts Journal*, vol. 64, no. 2: 30–40.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2009. *Credit Suisse Global Investment Returns Sourcebook*. 2009. Zurich, Switzerland: Credit Suisse Research Institute.

Taleb, Nassim N. 2007. *The Black Swan: The Impact of the Highly Improbable*. New York: Random House Inc.

PRACTICE PROBLEMS

1 An investor purchased 100 shares of a stock for \$34.50 per share at the beginning of the quarter. If the investor sold all of the shares for \$30.50 per share after receiving a \$51.55 dividend payment at the end of the quarter, the holding period return is *closest* to:

A −13.0%.

B −11.6%.

C −10.1%.

2 An analyst obtains the following annual rates of return for a mutual fund:

Year	Return (%)
2008	14
2009	−10
2010	−2

The fund's holding period return over the three-year period is *closest* to:

A 0.18%.

B 0.55%.

C 0.67%.

3 An analyst observes the following annual rates of return for a hedge fund:

Year	Return (%)
2008	22
2009	−25
2010	11

The hedge fund's annual geometric mean return is *closest* to:

A 0.52%.

B 1.02%.

C 2.67%.

4 Which of the following return calculating methods is *best* for evaluating the annualized returns of a buy-and-hold strategy of an investor who has made annual deposits to an account for each of the last five years?

A Geometric mean return.

B Arithmetic mean return.

C Money-weighted return.

5 An investor performs the following transactions on the shares of a firm.

- At t = 0, she purchases a share for \$1,000.
- At t = 1, she receives a dividend of \$25 and then purchases three additional shares for \$1,055 each.
- At t = 2, she receives a total dividend of \$100 and then sells the four shares for \$1,100 each.

The money-weighted rate of return is *closest* to:

A 4.5%.

B 6.9%.

C 7.3%.

6 A fund receives investments at the beginning of each year and generates returns as shown in the table.

Year of Investment	Assets Under Management at the beginning of each year	Return during Year of Investment
1	$1,000	15%
2	$4,000	14%
3	$45,000	−4%

Which return measure over the three-year period is negative?

A Geometric mean return

B Time-weighted rate of return

C Money-weighted rate of return

7 At the beginning of Year 1, a fund has $10 million under management; it earns a return of 14% for the year. The fund attracts another $100 million at the start of Year 2 and earns a return of 8% for that year. The money-weighted rate of return is *most likely*:

A less than the time-weighted rate of return.

B the same as the time-weighted rate of return.

C greater than the time-weighted rate of return.

8 An investor evaluating the returns of three recently formed exchange-traded funds gathers the following information:

ETF	Time Since Inception	Return Since Inception (%)
1	146 days	4.61
2	5 weeks	1.10
3	15 months	14.35

The ETF with the highest annualized rate of return is:

A ETF 1.

B ETF 2.

C ETF 3.

9 With respect to capital market theory, which of the following asset characteristics is *least likely* to impact the variance of an investor's equally weighted portfolio?

A Return on the asset.

B Standard deviation of the asset.

C Covariances of the asset with the other assets in the portfolio.

10 A portfolio manager creates the following portfolio:

Security	Security Weight (%)	Expected Standard Deviation (%)
1	30	20
2	70	12

If the correlation of returns between the two securities is 0.40, the expected standard deviation of the portfolio is *closest* to:

A 10.7%.

B 11.3%.

C 12.1%.

11 A portfolio manager creates the following portfolio:

Security	Security Weight (%)	Expected Standard Deviation (%)
1	30	20
2	70	12

If the covariance of returns between the two securities is −0.0240, the expected standard deviation of the portfolio is *closest* to:

A 2.4%.

B 7.5%.

C 9.2%.

The following information relates to Questions 12–13

A portfolio manager creates the following portfolio:

Security	Security Weight (%)	Expected Standard Deviation (%)
1	30	20
2	70	12

12 If the standard deviation of the portfolio is 14.40%, the correlation between the two securities is equal to:

A −1.0.

B 0.0.

C 1.0.

13 If the standard deviation of the portfolio is 14.40%, the covariance between the two securities is equal to:

A 0.0006.

B 0.0240.

C 1.0000.

The following information relates to Questions 14–17

An analyst observes the following historic geometric returns:

Asset Class	Geometric Return (%)
Equities	8.0
Corporate Bonds	6.5
Treasury bills	2.5
Inflation	2.1

14 The real rate of return for equities is *closest* to:

A 5.4%.

B 5.8%.

C 5.9%.

15 The real rate of return for corporate bonds is *closest* to:

A 4.3%.

B 4.4%.

C 4.5%.

16 The risk premium for equities is *closest* to:

A 5.4%.

B 5.5%.

C 5.6%.

17 The risk premium for corporate bonds is *closest* to:

A 3.5%.

B 3.9%.

C 4.0%.

---

18 With respect to trading costs, liquidity is *least likely* to impact the:

A stock price.

B bid–ask spreads.

C brokerage commissions.

19 Evidence of risk aversion is *best* illustrated by a risk–return relationship that is:

A negative.

B neutral.

C positive.

20 With respect to risk-averse investors, a risk-free asset will generate a numerical utility that is:

A the same for all individuals.

B positive for risk-averse investors.

C equal to zero for risk seeking investors.

21 With respect to utility theory, the most risk-averse investor will have an indifference curve with the:

A most convexity.

B smallest intercept value.

C greatest slope coefficient.

22 With respect to an investor's utility function expressed as: $U = E(r) - \frac{1}{2}A\sigma^2$, which of the following values for the measure for risk aversion has the *least* amount of risk aversion?

A −4.

B 0.

C 4.

The following information relates to Questions 23–26

A financial planner has created the following data to illustrate the application of utility theory to portfolio selection:

Investment	Expected Return (%)	Expected Standard Deviation (%)
1	18	2
2	19	8
3	20	15
4	18	30

23 A risk-neutral investor is *most likely* to choose:

A Investment 1.

B Investment 2.

C Investment 3.

24 If an investor's utility function is expressed as $U = E(r) - \frac{1}{2}A\sigma^2$ and the measure for risk aversion has a value of −2, the risk-seeking investor is *most likely* to choose:

A Investment 2.

B Investment 3.

C Investment 4.

25 If an investor's utility function is expressed as $U = E(r) - \frac{1}{2}A\sigma^2$ and the measure for risk aversion has a value of 2, the risk-averse investor is *most likely* to choose:

A Investment 1.

B Investment 2.

C Investment 3.

26 If an investor's utility function is expressed as $U = E(r) - \frac{1}{2}A\sigma^2$ and the measure for risk aversion has a value of 4, the risk-averse investor is *most likely* to choose:

A Investment 1.

B Investment 2.

C Investment 3.

---

27 With respect to the mean–variance portfolio theory, the capital allocation line, CAL, is the combination of the risk-free asset and a portfolio of all:

A risky assets.

B equity securities.

C feasible investments.

28 Two individual investors with different levels of risk aversion will have optimal portfolios that are:

A below the capital allocation line.

B on the capital allocation line.

C above the capital allocation line.

The following information relates to Questions 29–31

A portfolio manager creates the following portfolio:

Security	Expected Annual Return (%)	Expected Standard Deviation (%)
1	16	20
2	12	20

29 If the portfolio of the two securities has an expected return of 15%, the proportion invested in Security 1 is:

A 25%.

B 50%.

C 75%.

30 If the correlation of returns between the two securities is −0.15, the expected standard deviation of an equal-weighted portfolio is *closest* to:

A 13.04%.

B 13.60%.

C 13.87%.

31 If the two securities are uncorrelated, the expected standard deviation of an equal-weighted portfolio is *closest* to:

A 14.00%.

B 14.14%.

C 20.00%.

---

32 As the number of assets in an equally-weighted portfolio increases, the contribution of each individual asset's variance to the volatility of the portfolio:

A increases.

B decreases.

C remains the same.

33 With respect to an equally-weighted portfolio made up of a large number of assets, which of the following contributes the *most* to the volatility of the portfolio?

A Average variance of the individual assets.

B Standard deviation of the individual assets.

C Average covariance between all pairs of assets.

34 The correlation between assets in a two-asset portfolio increases during a market decline. If there is no change in the proportion of each asset held in the portfolio or the expected standard deviation of the individual assets, the volatility of the portfolio is *most likely* to:

A increase.

B decrease.

C remain the same.

The following information relates to Questions 35–37

An analyst has made the following return projections for each of three possible outcomes with an equal likelihood of occurrence:

Asset	Outcome 1 (%)	Outcome 2 (%)	Outcome 3 (%)	Expected Return (%)
1	12	0	6	6
2	12	6	0	6
3	0	6	12	6

35 Which pair of assets is perfectly negatively correlated?

A Asset 1 and Asset 2.

B Asset 1 and Asset 3.

C Asset 2 and Asset 3.

36 If the analyst constructs two-asset portfolios that are equally-weighted, which pair of assets has the *lowest* expected standard deviation?

A Asset 1 and Asset 2.

B Asset 1 and Asset 3.

C Asset 2 and Asset 3.

37 If the analyst constructs two-asset portfolios that are equally weighted, which pair of assets provides the *least* amount of risk reduction?

A Asset 1 and Asset 2.

B Asset 1 and Asset 3.

C Asset 2 and Asset 3.

---

38 Which of the following statements is *least* accurate? The efficient frontier is the set of all attainable risky assets with the:

A highest expected return for a given level of risk.

- **B** lowest amount of risk for a given level of return.
- **C** highest expected return relative to the risk-free rate.

39 The portfolio on the minimum-variance frontier with the lowest standard deviation is:

- **A** unattainable.
- **B** the optimal risky portfolio.
- **C** the global minimum-variance portfolio.

40 The set of portfolios on the minimum-variance frontier that dominates all sets of portfolios below the global minimum-variance portfolio is the:

- **A** capital allocation line.
- **B** Markowitz efficient frontier.
- **C** set of optimal risky portfolios.

41 The dominant capital allocation line is the combination of the risk-free asset and the:

- **A** optimal risky portfolio.
- **B** levered portfolio of risky assets.
- **C** global minimum-variance portfolio.

42 Compared to the efficient frontier of risky assets, the dominant capital allocation line has higher rates of return for levels of risk greater than the optimal risky portfolio because of the investor's ability to:

- **A** lend at the risk-free rate.
- **B** borrow at the risk-free rate.
- **C** purchase the risk-free asset.

43 With respect to the mean–variance theory, the optimal portfolio is determined by each individual investor's:

- **A** risk-free rate.
- **B** borrowing rate.
- **C** risk preference.

SOLUTIONS

1 C is correct. −10.1% is the holding period return, which is calculated as: (3,050 − 3,450 + 51.55)/3,450, which is comprised of a dividend yield of 1.49% = 51.55/(3,450) and a capital loss yield of −11.59% = −400/(3,450).

2 B is correct. [(1 + 0.14)(1 − 0.10)(1 − 0.02)] − 1 = 0.0055 = 0.55%.

3 A is correct. $[(1 + 0.22)(1 - 0.25)(1 + 0.11)]^{(1/3)} - 1 = 1.0157^{(1/3)} - 1 = 0.0052 =$ 0.52%

4 A is correct. The geometric mean return compounds the returns instead of the amount invested.

5 B is correct. Computation of the money-weighted return, *r*, requires finding the discount rate that sums the present value of cash flows to zero.

The first step is to group net cash flows by time. For this example, we have −\$1,000 for the $t = 0$ net cash flow, −\$3,140 = −\$3,165 + \$25 for the $t = 1$ net cash flow, and \$4,500 = \$4,400 + \$100 for the $t = 2$ net cash flow

Solving for *r*,

$$CF_0 = -1{,}000$$
$$CF_1 = -3{,}140$$
$$CF_2 = +4{,}500$$
$$\frac{CF_0}{(1+IRR)^0} + \frac{CF_1}{(1+IRR)^1} + \frac{CF_2}{(1+IRR)^2}$$
$$= \frac{-1{,}000}{1} + \frac{-3{,}140}{(1+IRR)^1} + \frac{4{,}500}{(1+IRR)^2} = 0$$

results in a value of $r = 6.91\%$

6 C is correct. The money-weighted rate of return considers both the timing and amounts of investments into the fund. To calculate the money-weighted rate of return, tabulate the annual returns and investment amounts to determine the cash flows

Year	1	2	3
Balance from previous year	0	\$1,150	\$4,560
New investment	\$1,000	\$2,850	\$40,440
Net balance at the beginning of year	\$1,000	\$4,000	\$45,000
Investment return for the year	15%	14%	−4%
Investment gain (loss)	\$150	\$560	−\$1,800
Balance at the end of year	\$1,150	\$4,560	\$43,200

$$CF_0 = -\$1{,}000,\ CF_1 = -\$2{,}850,\ CF_2 = -\$40{,}440,\ CF_3 = +\$43{,}200$$

Each cash inflow or outflow occurs at the end of each year. Thus, CF_0 refers to the cash flow at the end of Year 0 or beginning of Year 1, and CF_3 refers to the cash flow at end of Year 3 or beginning of Year 4. Because cash flows are being discounted to the present—that is, end of Year 0 or beginning of Year 1—the period of discounting CF_0 is zero whereas the period of discounting for CF_3 is 3 years.

Solving for *r*,

$$CF_0 = -1{,}000$$
$$CF_1 = -2{,}850$$
$$CF_2 = -40{,}440$$
$$CF_3 = +43{,}200$$

$$\frac{CF_0}{(1+IRR)^0} + \frac{CF_1}{(1+IRR)^1} + \frac{CF_2}{(1+IRR)^2} + \frac{CF_3}{(1+IRR)^3} = \frac{-1{,}000}{1} + \frac{-2{,}850}{(1+IRR)^1} + \frac{-40{,}440}{(1+IRR)^2} + \frac{43{,}200}{(1+IRR)^3} = 0$$

results in a value of $r = -2.22\%$

Note that B is incorrect because the time-weighted rate of return (TWR) of the fund is the same as the geometric mean return of the fund and is thus positive:

$$TWR = \sqrt[3]{(1.15)(1.14)(0.96)} - 1 = 7.97\%$$

7 A is correct. Computation of the money-weighted return, *r*, requires finding the discount rate that sums the present value of cash flows to zero. Because most of the investment came during Year 2, the measure will be biased toward the performance of Year 2. The cash flows are as follows:

$$CF_0 = -10$$
$$CF_1 = -100$$
$$CF_2 = +120.31$$

The terminal value is determined by summing the investment returns for each period [(10 × 1.14 × 1.08) + (100 × 1.08)]

$$\frac{CF_0}{(1+IRR)^0} + \frac{CF_1}{(1+IRR)^1} + \frac{CF_2}{(1+IRR)^2} = \frac{-10}{1} + \frac{-100}{(1+IRR)^1} + \frac{120.31}{(1+IRR)^2}$$

results in a value of $r = 8.53\%$

The time-weighted return of the fund is $= \sqrt[2]{(1.14)(1.08)} - 1 = 10.96\%$.

8 B is correct. The annualized rate of return for ETF 2 is $12.05\% = (1.0110^{52/5}) - 1$, which is greater than the annualized rate of ETF 1, $11.93\% = (1.0461^{365/146}) - 1$, and ETF 3, $11.32\% = (1.1435^{12/15}) - 1$. Despite having the lowest value for the periodic rate, ETF 2 has the highest annualized rate of return because of the reinvestment rate assumption and the compounding of the periodic rate.

9 A is correct. The asset's returns are not used to calculate the portfolio's variance [only the assets' weights, standard deviations (or variances), and covariances (or correlations) are used].

10 C is correct.

$$\sigma_{port} = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{1,2}\sigma_1\sigma_2}$$
$$= \sqrt{(0.3)^2(20\%)^2 + (0.7)^2(12\%)^2 + 2(0.3)(0.7)(0.40)(20\%)(12\%)}$$
$$= (0.3600\% + 0.7056\% + 0.4032\%)^{0.5} = (1.4688\%)^{0.5} = 12.11\%$$

11 A is correct.

$$\sigma_{port} = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2Cov(R_1R_2)}$$
$$= \sqrt{(0.3)^2(20\%)^2 + (0.7)^2(12\%)^2 + 2(0.3)(0.7)(-0.0240)}$$
$$= (0.3600\% + 0.7056\% - 1.008\%)^{0.5} = (0.0576\%)^{0.5} = 2.40\%$$

12 C is correct. A portfolio standard deviation of 14.40% is the weighted average, which is possible only if the correlation between the securities is equal to 1.0.

13 B is correct. A portfolio standard deviation of 14.40% is the weighted average, which is possible only if the correlation between the securities is equal to 1.0. If the correlation coefficient is equal to 1.0, then the covariance must equal 0.0240, calculated as: $Cov(R_1,R_2) = \rho_{12}\sigma_1\sigma_2$ = (1.0)(20%)(12%) = 2.40% = 0.0240.

14 B is correct. (1 + 0.080)/(1 + 0.0210) − 1 = 5.8%

15 A is correct. (1 + 0.065)/(1 + 0.0210) − 1 = 4.3%

16 A is correct. (1 + 0.080)/(1 + 0.0250) − 1 = 5.4%

17 B is correct. (1 + 0.0650)/(1 + 0.0250) − 1 = 3.9%

18 C is correct. Brokerage commissions are negotiated with the brokerage firm. A security's liquidity impacts the operational efficiency of trading costs. Specifically, liquidity impacts the bid–ask spread and can impact the stock price (if the ability to sell the stock is impaired by the uncertainty associated with being able to sell the stock).

19 C is correct. Historical data over long periods of time indicate that there exists a positive risk–return relationship, which is a reflection of an investor's risk aversion.

20 A is correct. A risk-free asset has a variance of zero and is not dependent on whether the investor is risk neutral, risk seeking or risk averse. That is, given that the utility function of an investment is expressed as $U = E(r) - \frac{1}{2}A\sigma^2$, where A is the measure of risk aversion, then the sign of A is irrelevant if the variance is zero (like that of a risk-free asset).

21 C is correct. The most risk-averse investor has the indifference curve with the greatest slope.

22 A is correct. A negative value in the given utility function indicates that the investor is a risk seeker.

23 C is correct. Investment 3 has the highest rate of return. Risk is irrelevant to a risk-neutral investor, who would have a measure of risk aversion equal to 0. Given the utility function, the risk-neutral investor would obtain the greatest amount of utility from Investment 3.

Investment	Expected Return (%)	Expected Standard Deviation (%)	Utility $A = 0$
1	18	2	0.1800
2	19	8	0.1900
3	20	15	0.2000
4	18	30	0.1800

24 C is correct. Investment 4 provides the highest utility value (0.2700) for a risk-seeking investor, who has a measure of risk aversion equal to −2.

Investment	Expected Return (%)	Expected Standard Deviation (%)	Utility $A = -2$
1	18	2	0.1804
2	19	8	0.1964
3	20	15	0.2225
4	18	30	0.2700

25 B is correct. Investment 2 provides the highest utility value (0.1836) for a risk-averse investor who has a measure of risk aversion equal to 2.

Investment	Expected Return (%)	Expected Standard Deviation (%)	Utility $A = 2$
1	18	2	0.1796
2	19	8	0.1836
3	20	15	0.1775
4	18	30	0.0900

26 A is correct. Investment 1 provides the highest utility value (0.1792) for a risk-averse investor who has a measure of risk aversion equal to 4.

Investment	Expected Return (%)	Expected Standard Deviation (%)	Utility $A = 4$
1	18	2	0.1792
2	19	8	0.1772
3	20	15	0.1550
4	18	30	0.0000

27 A is correct. The CAL is the combination of the risk-free asset with zero risk and the portfolio of all risky assets that provides for the set of feasible investments. Allowing for borrowing at the risk-free rate and investing in the portfolio of all risky assets provides for attainable portfolios that dominate risky assets below the CAL.

28 B is correct. The CAL represents the set of all feasible investments. Each investor's indifference curve determines the optimal combination of the risk-free asset and the portfolio of all risky assets, which must lie on the CAL.

29 C is correct.

$$R_p = w_1 \times R_1 + (1 - w_1) \times R_2$$
$$R_p = w_1 \times 16\% + (1 - w_1) \times 12\%$$
$$15\% = 0.75(16\%) + 0.25(12\%)$$

30 A is correct.

$$\begin{aligned}\sigma_{port} &= \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{1,2}\sigma_1\sigma_2} \\ &= \sqrt{(0.5)^2(20\%)^2 + (0.5)^2(20\%)^2 + 2(0.5)(0.5)(-0.15)(20\%)(20\%)} \\ &= (1.0000\% + 1.0000\% - 0.3000\%)^{0.5} = (1.7000\%)^{0.5} = 13.04\%\end{aligned}$$

31 B is correct.

$$\begin{aligned}\sigma_{port} &= \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{1,2}\sigma_1\sigma_2} \\ &= \sqrt{(0.5)^2(20\%)^2 + (0.5)^2\left(20\%^2\right) + 2(0.5)(0.5)(0.00)(20\%)(20\%)} \\ &= (1.0000\% + 1.0000\% + 0.0000\%)^{0.5} = (2.0000\%)^{0.5} = 14.14\%\end{aligned}$$

32 B is correct. The contribution of each individual asset's variance (or standard deviation) to the portfolio's volatility decreases as the number of assets in the equally weighted portfolio increases. The contribution of the co-movement measures between the assets increases (i.e., covariance and correlation) as the number of assets in the equally weighted portfolio increases. The following equation for the variance of an equally weighted portfolio illustrates these points: $\sigma_p^2 = \frac{\bar{\sigma}^2}{N} + \frac{N-1}{N}\overline{COV} = \frac{\bar{\sigma}^2}{N} + \frac{N-1}{N}\bar{\rho}\bar{\sigma}^2$.

33 C is correct. The co-movement measures between the assets increases (i.e., covariance and correlation) as the number of assets in the equally weighted portfolio increases. The contribution of each individual asset's variance (or standard deviation) to the portfolio's volatility decreases as the number of assets in the equally weighted portfolio increases. The following equation for the variance of an equally weighted portfolio illustrates these points:

$$\sigma_p^2 = \frac{\bar{\sigma}^2}{N} + \frac{N-1}{N}\overline{COV} = \frac{\bar{\sigma}^2}{N} + \frac{N-1}{N}\bar{\rho}\bar{\sigma}^2$$

34 A is correct. Higher correlations will produce less diversification benefits provided that the other components of the portfolio standard deviation do not change (i.e., the weights and standard deviations of the individual assets).

35 C is correct. Asset 2 and Asset 3 have returns that are the same for Outcome 2, but the exact opposite returns for Outcome 1 and Outcome 3; therefore, because they move in opposite directions at the same magnitude, they are perfectly negatively correlated.

36 C is correct. An equally weighted portfolio of Asset 2 and Asset 3 will have the lowest portfolio standard deviation, because for each outcome, the portfolio has the same expected return (they are perfectly negatively correlated).

37 A is correct. An equally weighted portfolio of Asset 1 and Asset 2 has the highest level of volatility of the three pairs. All three pairs have the same expected return; however, the portfolio of Asset 1 and Asset 2 provides the least amount of risk reduction.

38 C is correct. The efficient frontier does not account for the risk-free rate. The efficient frontier is the set of all attainable risky assets with the highest expected return for a given level of risk or the lowest amount of risk for a given level of return.

39 C is correct. The global minimum-variance portfolio is the portfolio on the minimum-variance frontier with the lowest standard deviation. Although the portfolio is attainable, when the risk-free asset is considered, the global minimum-variance portfolio is not the optimal risky portfolio.

40 B is correct. The Markowitz efficient frontier has higher rates of return for a given level of risk. With respect to the minimum-variance portfolio, the Markowitz efficient frontier is the set of portfolios above the global minimum-variance portfolio that dominates the portfolios below the global minimum-variance portfolio.

41 A is correct. The use of leverage and the combination of a risk-free asset and the optimal risky asset will dominate the efficient frontier of risky assets (the Markowitz efficient frontier).

42 B is correct. The CAL dominates the efficient frontier at all points except for the optimal risky portfolio. The ability of the investor to purchase additional amounts of the optimal risky portfolio by borrowing (i.e., buying on margin) at the risk-free rate makes higher rates of return for levels of risk greater than the optimal risky asset possible.

43 C is correct. Each individual investor's optimal mix of the risk-free asset and the optimal risky asset is determined by the investor's risk preference.

READING

50

Portfolio Risk and Return: Part II

by Vijay Singal, PhD, CFA

Vijay Singal, PhD, CFA, is at Virginia Tech (USA).

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**a.** describe the implications of combining a risk-free asset with a portfolio of risky assets;
☐	**b.** explain the capital allocation line (CAL) and the capital market line (CML);
☐	**c.** explain systematic and nonsystematic risk, including why an investor should not expect to receive additional return for bearing nonsystematic risk;
☐	**d.** explain return generating models (including the market model) and their uses;
☐	**e.** calculate and interpret beta;
☐	**f.** explain the capital asset pricing model (CAPM), including its assumptions, and the security market line (SML);
☐	**g.** calculate and interpret the expected return of an asset using the CAPM;
☐	**h.** describe and demonstrate applications of the CAPM and the SML;
☐	**i.** calculate and interpret the Sharpe ratio, Treynor ratio, M^2, and Jensen's alpha.

INTRODUCTION

1

Our objective in this reading is to identify the optimal risky portfolio for all investors by using the capital asset pricing model (CAPM). The foundation of this reading is the computation of risk and return of a portfolio and the role that correlation plays in diversifying portfolio risk and arriving at the efficient frontier. The efficient frontier and the capital allocation line consist of portfolios that are generally acceptable to all investors. By combining an investor's individual indifference curves with the market-determined capital allocation line, we are able to illustrate that the only optimal risky portfolio for an investor is the portfolio of all risky assets (i.e., the market).

Additionally, we discuss the capital market line, a special case of the capital allocation line that is used for passive investor portfolios. We also differentiate between systematic and nonsystematic risk, and explain why investors are compensated for bearing systematic risk but receive no compensation for bearing nonsystematic risk. We discuss in detail the CAPM, which is a simple model for estimating asset returns based only on the asset's systematic risk. Finally, we illustrate how the CAPM allows security selection to build an optimal portfolio for an investor by changing the asset mix beyond a passive market portfolio.

The reading is organized as follows. In Sections 2–4, we discuss the consequences of combining a risk-free asset with the market portfolio and provide an interpretation of the capital market line. Sections 5–7 decompose total risk into systematic and nonsystematic risk and discuss the characteristics of and differences between the two kinds of risk. We also introduce return-generating models, including the single-index model, and illustrate the calculation of beta. In Sections 8 and 9, we introduce the capital asset pricing model and the security market line. Our focus on the CAPM does not suggest that the CAPM is the only viable asset pricing model. Although the CAPM is an excellent starting point, more advanced readings expand on these discussions and extend the analysis to other models that account for multiple explanatory factors. Section 10 covers several post-CAPM developments in theory. Section 11 covers measures for evaluating the performance of a portfolio which take account of risk. Section 12 covers some applications of the CAPM in portfolio construction. A summary and practice problems conclude the reading.

2 CAPITAL MARKET THEORY: RISK-FREE AND RISKY ASSETS

a describe the implications of combining a risk-free asset with a portfolio of risky assets;

b explain the capital allocation line (CAL) and the capital market line (CML);

You have learned how to combine a risk-free asset with one risky asset and with many risky assets to create a capital allocation line. In this section, we will expand our discussion of multiple risky assets and consider a special case of the capital allocation line, called the capital market line. While discussing the capital market line, we will define the market and its role in passive portfolio management. Using these concepts, we will illustrate how leveraged portfolios can enhance both risk and return.

2.1 Portfolio of Risk-Free and Risky Assets

Although investors desire an asset that produces the highest return and carries the lowest risk, such an asset does not exist. As the risk–return capital market theory illustrates, one must assume higher risk in order to earn a higher return. We can improve an investor's portfolio, however, by expanding the opportunity set of risky assets because this allows the investor to choose a superior mix of assets.

Similarly, an investor's portfolio improves if a risk-free asset is added to the mix. In other words, a combination of the risk-free asset and a risky asset can result in a better risk–return trade-off than an investment in only one type of asset because the risk-free asset has zero correlation with the risky asset. The combination is called the **capital allocation line** (and is depicted in Exhibit 2). Superimposing an investor's indifference curves on the capital allocation line will lead to the optimal investor portfolio.

Investors with different levels of risk aversion will choose different portfolios. Highly risk-averse investors choose to invest most of their wealth in the risk-free asset and earn low returns because they are not willing to assume higher levels of risk. Less risk-averse investors, in contrast, invest more of their wealth in the risky asset, which is expected to yield a higher return. Obviously, the higher return cannot come without higher risk, but the less risk-averse investor is willing to accept the additional risk.

2.1.1 *Combining a Risk-Free Asset with a Portfolio of Risky Assets*

We can extend the analysis of one risky asset to a portfolio of risky assets. For convenience, assume that the portfolio contains all available risky assets (*N*), although an investor may not wish to include all of these assets in the portfolio because of the investor's specific preferences. If an asset is not included in the portfolio, its weight will be zero. The risk–return characteristics of a portfolio of *N* risky assets are given by the following equations:

$$E\left(R_p\right) = \sum_{i=1}^{N} w_i E\left(R_i\right)$$

$$\sigma_p^2 = \left(\sum_{i=1,j=1}^{N} w_i w_j \text{Cov}\left(i,j\right)\right), \text{ and } \sum_{i=1}^{N} w_i = 1$$

The expected return on the portfolio, $E(R_p)$, is the weighted average of the expected returns of individual assets, where w_i is the fractional weight in asset *i* and R_i is the expected return of asset *i*. The risk of the portfolio (σ_p), however, depends on the weights of the individual assets, the risk of the individual assets, and their interrelationships. The **covariance** between assets *i* and *j*, Cov(*i*,*j*), is a statistical measure of the interrelationship between each pair of assets in the portfolio and can be expressed as follows, where ρ_{ij} is the **correlation** between assets *i* and *j* and σ_i is the risk of asset *i*:

$$\text{Cov}(i,j) = \rho_{ij}\sigma_i\sigma_j$$

Note from the equation below that the correlation of an asset with itself is 1; therefore:

$$\text{Cov}(i,i) = \rho_{ii}\sigma_i\sigma_i = \sigma_i^2$$

By substituting the above expressions for covariance, we can rewrite the portfolio variance equation as

$$\sigma_p^2 = \left(\sum_{i=1}^{N} w_i^2\sigma_i^2 + \sum_{i,j=1,i\neq j}^{N} w_i w_j \rho_{ij}\sigma_i\sigma_j\right)$$

The suggestion that portfolios have lower risk than the assets they contain may seem counterintuitive. These portfolios can be constructed, however, as long as the assets in the portfolio are not perfectly correlated. As an illustration of the effect of asset weights on portfolio characteristics, consider a simple two-asset portfolio with zero weights in all other assets. Assume that Asset 1 has a return of 10 percent and a standard deviation (risk) of 20 percent. Asset 2 has a return of 5 percent and a standard deviation (risk) of 10 percent. Furthermore, the correlation between the two assets is zero. Exhibit 1 shows risks and returns for Portfolio X with a weight of 25 percent in Asset 1 and 75 percent in Asset 2, Portfolio Y with a weight of 50 percent in each of the two assets, and Portfolio Z with a weight of 75 percent in Asset 1 and 25 percent in Asset 2.

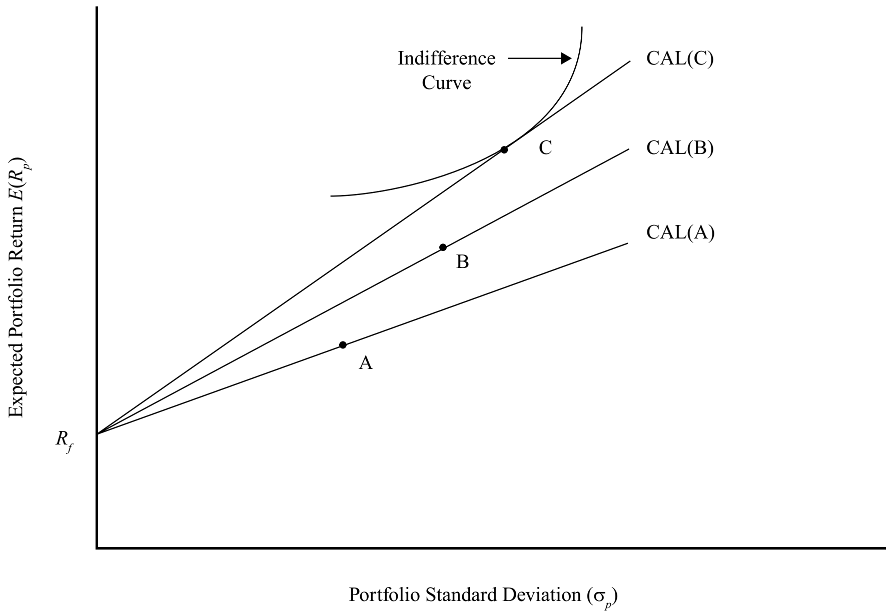
Exhibit 1 Portfolio Risk and Return

Portfolio	Weight in Asset 1 (%)	Weight in Asset 2 (%)	Portfolio Return (%)	Portfolio Standard Deviation (%)
X	25.0	75.0	6.25	9.01
Y	50.0	50.0	7.50	11.18
Z	75.0	25.0	8.75	15.21
Return	10.0	5.0		
Standard deviation	20.0	10.0		
Correlation between Assets 1 and 2		0.0		

From this example we observe that the three portfolios are quite different in terms of their risk and return. Portfolio X has a 6.25 percent return and only 9.01 percent standard deviation, whereas the standard deviation of Portfolio Z is more than two-thirds higher (15.21 percent), although the return is only slightly more than one-third higher (8.75 percent). These portfolios may become even more dissimilar as other assets are added to the mix.

Consider three portfolios of risky assets, A, B, and C, as in Exhibit 2, that may have been presented to a representative investor by three different investment advisers. Each portfolio is combined with the risk-free asset to create three capital allocation lines, CAL(A), CAL(B), and CAL(C). The exhibit shows that Portfolio C is superior to the other two portfolios because it has a greater expected return for any given level of risk. As a result, an investor will choose the portfolio that lies on the capital allocation line for Portfolio C. The combination of the risk-free asset and the risky Portfolio C that is selected for an investor depends on the investor's degree of risk aversion.

Exhibit 2 Risk-Free Asset and Portfolio of Risky Assets



2.1.2 *Does a Unique Optimal Risky Portfolio Exist?*

We assume that all investors have the same economic expectation and thus have the same expectations of prices, cash flows, and other investment characteristics. This assumption is referred to as **homogeneity of expectations**. Given these investment characteristics, everyone goes through the same calculations and should arrive at the same optimal risky portfolio. Therefore, assuming homogeneous expectations, only one optimal portfolio exists. If investors have different expectations, however, they might arrive at different optimal risky portfolios. To illustrate, we begin with an expression for the price of an asset:

$$P = \sum_{t=0}^{T} \frac{\mathrm{CF}_t}{\left(1 + r_t\right)^t}$$

where CF_t is the cash flow at the end of period t and r_t is the discount rate or the required rate of return for that asset for period t. Period t refers to all periods beginning from now until the asset ceases to exist at the end of time T. Because the current time is the end of period 0, which is the same as the beginning of period 1, there are (T + 1) cash flows and (T + 1) required rates of return. These conditions are based on the assumption that a cash flow, such as an initial investment, can occur now (t = 0). Ordinarily, however, CF_0 is zero.

We use the formula for the price of an asset to estimate the intrinsic value of an asset. Assume that the asset we are valuing is a share of Siemens AG which trades on Xetra. In the case of corporate stock, there is no expiration date, so T could be extremely large, meaning we will need to estimate a large number of cash flows and rates of return. Fortunately, the denominator reduces the importance of distant cash flows, so it may be sufficient to estimate, say, 20 annual cash flows and 20 rates of returns. How much will Siemens earn next year and the year after next? What will the product markets Siemens operates in look like in five years' time? Different analysts and investors will have their own estimates that may be quite different from one another. Also, as we delve further into the future, more serious issues in estimating future revenue, expenses, and growth rates arise. Therefore, to assume that cash flow estimates for Siemens will vary among these investors is reasonable. In addition to the numerator (cash flows), it is also necessary to estimate the denominator, the required rates of return. We know that riskier companies will require higher returns because risk and return are positively correlated. Siemens stock is riskier than a risk-free asset, but by how much? And what should the compensation for that additional risk be? Again, it is evident that different analysts will view the riskiness of Siemens differently and, therefore, arrive at different required rates of return.

Siemens closed at €111.84 on Xetra on 31 August 2018. The traded price represents the value that a marginal investor attaches to a share of Siemens, say, corresponding to Analyst A's expectation. Analyst B may think that the price should be €95, however, and Analyst C may think that the price should be €125. Given a price of €111.84, the expected returns of Siemens are quite different for the three analysts. Analyst B, who believes the price should be €95, concludes that Siemens is overvalued and may assign a weight of zero to Siemens in the recommended portfolio even though the market capitalization of Siemens was in excess of €100 billion as of the date of the quotation. In contrast, Analyst C, with a valuation of €125, thinks Siemens is undervalued and may significantly overweight Siemens in a portfolio.

Our discussion illustrates that analysts can arrive at different valuations that necessitate the assignment of different asset weights in a portfolio. Given the existence of many asset classes and numerous assets in each asset class, one can visualize that each investor will have his or her own optimal risky portfolio depending on his or her assumptions underlying the valuation computations. Therefore, market participants will have their own and possibly different optimal risky portfolios.

If investors have different valuations of assets, then the construction of a unique optimal risky portfolio is not possible. If we make a simplifying assumption of homogeneity in investor expectations, we will have a single optimal risky portfolio as previously mentioned. Even if investors have different expectations, market prices are a proxy of what the marginal, informed investor expects, and the market portfolio becomes the base case, the benchmark, or the reference portfolio that other portfolios can be judged against. For Siemens, the market price was €111.84 per share and the market capitalization was about €108 billion. In constructing the market portfolio, Siemens's weight in the market portfolio will be equal to its market value divided by the value of all other assets included in the market portfolio.

3 CAPITAL MARKET THEORY: THE CAPITAL MARKET LINE

b explain the capital allocation line (CAL) and the capital market line (CML);

In the previous section, we discussed how the risk-free asset could be combined with a risky portfolio to create a capital allocation line (CAL). In this section, we discuss a specific CAL that uses the market portfolio as the optimal risky portfolio and is known as the capital market line. We also discuss the significance of the market portfolio and applications of the capital market line (CML).

3.1 Passive and Active Portfolios

In the above subsection, we hypothesized three possible valuations for each share of Siemens: €95, €111.84, and €125. Which one is correct?

If the market is an **informationally efficient market**, the price in the market, €111.84, is an unbiased estimate of all future discounted cash flows (recall the formula for the price of an asset). In other words, the price aggregates and reflects all information that is publicly available, and investors cannot expect to earn a return that is greater than the required rate of return for that asset. If, however, the price reflects all publicly available information and there is no way to outperform the market, then there is little point in investing time and money in evaluating Siemens to arrive at your price using your own estimates of cash flows and rates of return.

In that case, a simple and convenient approach to investing is to rely on the prices set by the market. Portfolios that are based on the assumption of unbiased market prices are referred to as passive portfolios. Passive portfolios most commonly replicate and track market indexes, which are passively constructed on the basis of market prices and market capitalizations. Examples of market indexes are the S&P 500 Index, the Nikkei 300, and the CAC 40. Passive portfolios based on market indexes are called index funds and generally have low costs because no significant effort is expended in valuing securities that are included in an index.

In contrast to passive investors' reliance on market prices and index funds, active investors may not rely on market valuations. They have more confidence in their own ability to estimate cash flows, growth rates, and discount rates. Based on these estimates, they value assets and determine whether an asset is fairly valued. In an actively managed portfolio, assets that are undervalued, or have a chance of offering above-normal returns, will have a positive weight (i.e., overweight compared to the market weight in the benchmark index), whereas other assets will have a zero weight, or even a negative weight if short selling is permitted (i.e., some assets will be underweighted compared with the market weight in the benchmark index). (**Short selling** is a transaction in which borrowed securities are sold with the intention to repurchase them at a lower price at a later date and return them to the lender.) This style of investing is called active investment management, and the portfolios are referred to as active portfolios. Most open-end mutual funds and hedge funds practice active investment management, and most analysts believe that active investing adds value. Whether these analysts are right or wrong is the subject of continuing debate.

3.2 What Is the "Market"?

In the previous discussion, we referred to the "market" on numerous occasions without actually defining the market. The optimal risky portfolio and the capital market line depend on the definition of the market. So what is the market?

Theoretically, the market includes all risky assets or anything that has value, which includes stocks, bonds, real estate, and even human capital. Not all assets are tradable, however, and not all tradable assets are investable. For example, the Taj Mahal in India is an asset but is not a tradable asset. Similarly, human capital is an asset that is not tradable. Moreover, assets may be tradable but not investable because of restrictions placed on certain kinds of investors. For example, all stocks listed on the Shanghai Stock Exchange are tradable. However, whereas Class A shares are listed in RMB and open to domestic investors and qualified foreign investors, Class B shares are listed in USD and open to foreign investors and domestic investors holding foreign currency dealing accounts.

If we consider all stocks, bonds, real estate assets, commodities, etc., probably hundreds of thousands of assets are tradable and investable. The "market" should contain as many assets as possible; we emphasize the word "possible" because it is not practical to include all assets in a single risky portfolio. Even though advancements in technology and interconnected markets have made it much easier to span the major equity markets, we are still not able to easily invest in other kinds of assets like bonds and real estate except in the most developed countries.

For the rest of this reading, we will define the "market" quite narrowly because it is practical and convenient to do so. Typically, a local or regional stock market index is used as a proxy for the market because of active trading in stocks and because a local or regional market is most visible to the local investors. For our purposes, we will use the S&P 500 Index as the market's proxy. The S&P 500 is commonly used by analysts as a benchmark for market performance throughout the United States. It contains 500 of the largest stocks that are domiciled in the United States, and these stocks are weighted by their market capitalization (price times the number of outstanding shares).

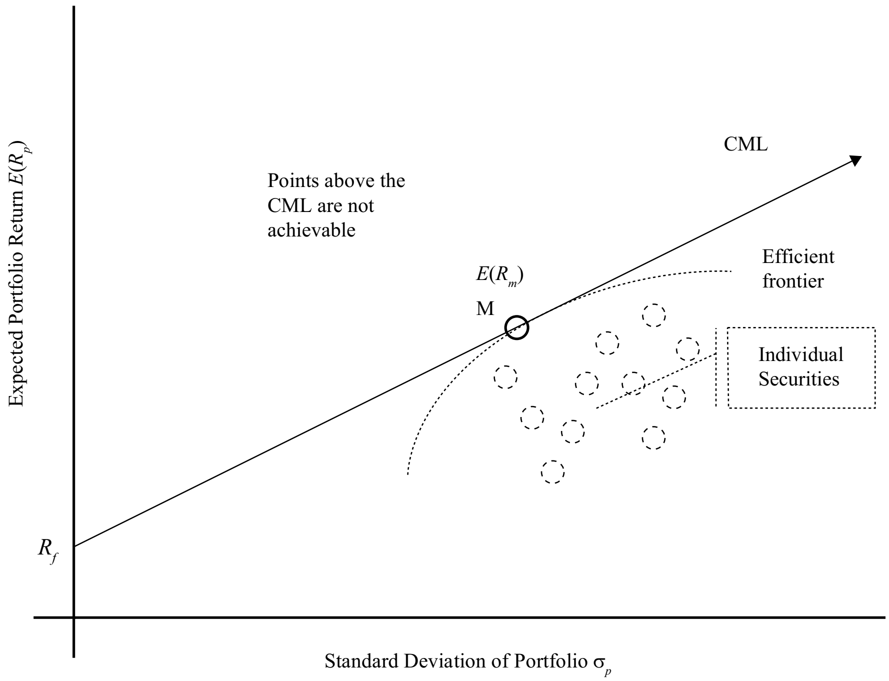
As of mid-2018, the stocks in the S&P 500 account for approximately 80 percent of the total equity market capitalization in the United States, and because the US stock markets represent about 40 percent of the world markets, the S&P 500 represents roughly 32 percent of worldwide publicly traded equity. Our definition of the market does not include non-US stock markets, bond markets, real estate, and many other asset classes, and therefore, "market" return and the "market" risk premium refer to US equity return and the US equity risk premium, respectively. The use of this proxy, however, is sufficient for our discussion, and is relatively easy to expand to include other tradable assets.

3.3 The Capital Market Line (CML)

A capital allocation line includes all possible combinations of the risk-free asset and an investor's optimal risky portfolio. The **capital market line** is a special case of the capital allocation line, where the risky portfolio is the market portfolio. The risk-free asset is a debt security with no default risk, no inflation risk, no liquidity risk, no interest rate risk, and no risk of any other kind. US Treasury bills are usually used as a proxy of the risk-free return, R_f.

The S&P 500 is a proxy of the market portfolio, which is the optimal risky portfolio. Therefore, the expected return on the risky portfolio is the expected market return, expressed as $E(R_m)$. The capital market line is shown in Exhibit 3, where the standard deviation (σ_p), or total risk, is on the x-axis and expected portfolio return, $E(R_p)$, is on the y-axis. Graphically, the market portfolio is the point on the Markowitz efficient frontier where a line from the risk-free asset is tangent to the Markowitz efficient frontier. All points on the interior of the Markowitz efficient frontier are inefficient portfolios in that they provide the same level of return with a higher level of risk or a lower level of return with the same amount of risk. When plotted together, the point at which the CML is tangent to the Markowitz efficient frontier is the optimal combination of risky assets, on the basis of market prices and market capitalizations. The optimal risky portfolio is the market portfolio.

Exhibit 3 Capital Market Line



The CML's intercept on the y-axis is the risk-free return (R_f) because that is the return associated with zero risk. The CML passes through the point represented by the market return, $E(R_m)$. With respect to capital market theory, any point above the CML is not achievable and any point below the CML is dominated by and inferior to any point on the CML.

Note that we identify the CML and CAL as lines even though they are a combination of two assets. Unlike a combination of two risky assets, which is usually not a straight line, a combination of the risk-free asset and a risky portfolio is a straight line, as illustrated below by computing the combination's risk and return.

Risk and return characteristics of the portfolio represented by the CML can be computed by using the return and risk expressions for a two-asset portfolio:

$$E(R_p) = w_1 R_f + (1 - w_1)E(R_m),$$

and

$$\sigma_p = \sqrt{w_1^2\sigma_f^2 + (1 - w_1)^2\sigma_m^2 + 2w_1(1 - w_1)\text{Cov}(R_f, R_m)}$$

The proportion invested in the risk-free asset is given by w_1, and the balance is invested in the market portfolio, $(1 - w_1)$. The risk of the risk-free asset is given by σ_f, the risk of the market is given by σ_m, the risk of the portfolio is given by σ_p, and the covariance between the risk-free asset and the market portfolio is represented by $\text{Cov}(R_f, R_m)$.

By definition, the standard deviation of the risk-free asset is zero. Because its risk is zero, the risk-free asset does not co-vary or move with any other asset. Therefore, its covariance with all other assets, including the market portfolio, is zero, making the first and third terms under the square root sign zero. As a result, the portfolio return and portfolio standard deviation can be simplified and rewritten as:

$$E(R_p) = w_1R_f + (1 - w_1)E(R_m),$$

and

$$\sigma_p = (1 - w_1)\sigma_m$$

By substitution, we can express $E(R_p)$ in terms of σ_p. Substituting for w_1, we get:

$$E(R_p) = R_f + \left(\frac{E(R_m) - R_f}{\sigma_m}\right) \times \sigma_p$$

Note that the expression is in the form of a line, $y = a + bx$. The y-intercept is the risk-free rate, and the slope of the line referred to as the market price of risk is $[E(R_m) - R_f]/\sigma_m$. The CML has a positive slope because the market's risky return is larger than the risk-free return. As the amount of the total investment devoted to the market increases—that is, as we move up the line—both standard deviation (risk) and expected return increase.

EXAMPLE 1

Risk and Return on the CML

Mr. Miles is a first time investor and wants to build a portfolio using only US T-bills and an index fund that closely tracks the S&P 500 Index. The T-bills have a return of 5 percent. The S&P 500 has a standard deviation of 20 percent and an expected return of 15 percent.

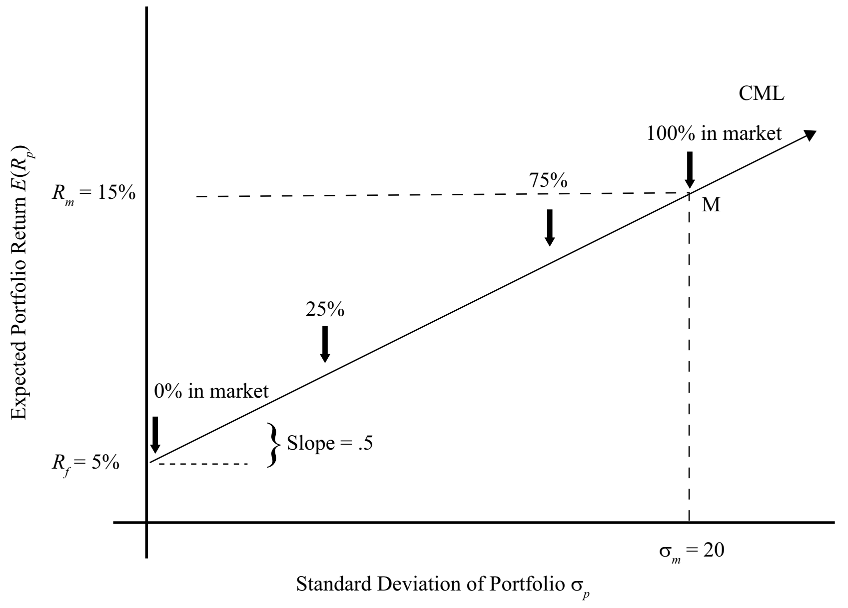
1. Draw the CML and mark the points where the investment in the market is 0 percent, 25 percent, 75 percent, and 100 percent.
2. Mr. Miles is also interested in determining the exact risk and return at each point.

Solution to 1:

We calculate the equation for the CML as $E(R_p) = 5\% + 0.50 \times \sigma_p$ by substituting the given information into the general CML equation. The intercept of the line is 5 percent, and its slope is 0.50. We can draw the CML by arbitrarily taking any two points on the line that satisfy the above equation.

Alternatively, the CML can be drawn by connecting the risk-free return of 5 percent on the y-axis with the market portfolio at (20 percent, 15 percent). The CML is shown in Exhibit 4.

Exhibit 4 Risk and Return on the CML



Solution to 2:

Return with 0 percent invested in the market = 5 percent, which is the risk-free return.

Standard deviation with 0 percent invested in the market = 0 percent because T-bills are not risky.

Return with 25 percent invested in the market = (0.75 × 5%) + (0.25 × 15%) = 7.5%.

Standard deviation with 25 percent invested in the market = 0.25 × 20% = 5%.

Return with 75 percent invested in the market = (0.25 × 5%) + (0.75 × 15%) = 12.50%.

Standard deviation with 75 percent invested in the market = 0.75 × 20% = 15%.

Return with 100 percent invested in the market = 15 percent, which is the return on the S&P 500.

Standard deviation with 100 percent invested in the market = 20 percent, which is the risk of the S&P 500.

4 CAPITAL MARKET THEORY: CML - LEVERAGED PORTFOLIOS

b explain the capital allocation line (CAL) and the capital market line (CML);

In the previous example, Mr. Miles evaluated an investment of between 0 percent and 100 percent in the market and the balance in T-bills. The line connecting R_f and M (market portfolio) in Exhibit 4 illustrates these portfolios with their respective levels of investment. At R_f, an investor is investing all of his or her wealth into risk-free securities, which is equivalent to lending 100 percent at the risk-free rate. At Point *M* he or she is holding the market portfolio and not lending any money at the risk-free rate. The combinations of the risk-free asset and the market portfolio, which may be achieved by the points between these two limits, are termed "lending" portfolios. In effect, the investor is lending part of his or her wealth at the risk-free rate.

If Mr. Miles is willing to take more risk, he may be able to move to the right of the market portfolio (Point *M* in Exhibit 4) by borrowing money and purchasing more of Portfolio M. Assume that he is able to borrow money at the same risk-free rate of interest, R_f, at which he can invest. He can then supplement his available wealth with borrowed money and construct a borrowing portfolio. If the straight line joining R_f and *M* is extended to the right of Point *M*, this extended section of the line represents borrowing portfolios. As one moves further to the right of Point *M*, an increasing amount of borrowed money is being invested in the market. This means that there is *negative* investment in the risk-free asset, which is referred to as a *leveraged position* in the risky portfolio. The particular point chosen on the CML will depend on the individual's utility function, which, in turn, will be determined by his risk and return preferences.

EXAMPLE 2

Risk and Return of a Leveraged Portfolio with Equal Lending and Borrowing Rates

Mr. Miles decides to set aside a small part of his wealth for investment in a portfolio that has greater risk than his previous investments because he anticipates that the overall market will generate attractive returns in the future. He assumes that he can borrow money at 5 percent and achieve the same return on the S&P 500 as before: an expected return of 15 percent with a standard deviation of 20 percent.

Calculate his expected risk and return if he borrows 25 percent, 50 percent, and 100 percent of his initial investment amount.

Solution:

The leveraged portfolio's standard deviation and return can be calculated in the same manner as before with the following equations:

$$E(R_p) = w_1 R_f + (1 - w_1)E(R_m)$$

and

$$\sigma_p = (1 - w_1)\sigma_m$$

The proportion invested in T-bills becomes negative instead of positive because Mr. Miles is borrowing money. If 25 percent of the initial investment is borrowed, $w_1 = -0.25$, and $(1 - w_1) = 1.25$, etc.

Return with $w_1 = -0.25 = (-0.25 \times 5\%) + (1.25 \times 15\%) = 17.5\%$.

Standard deviation with $w_1 = -0.25 = 1.25 \times 20\% = 25\%$.

Return with $w_1 = -0.50 = (-0.50 \times 5\%) + (1.50 \times 15\%) = 20.0\%$.

Standard deviation with $w_1 = -0.50 = 1.50 \times 20\% = 30\%$.

Return with $w_1 = -1.00 = (-1.00 \times 5\%) + (2.00 \times 15\%) = 25.0\%$.

Standard deviation with $w_1 = -1.00 = 2.00 \times 20\% = 40\%$.

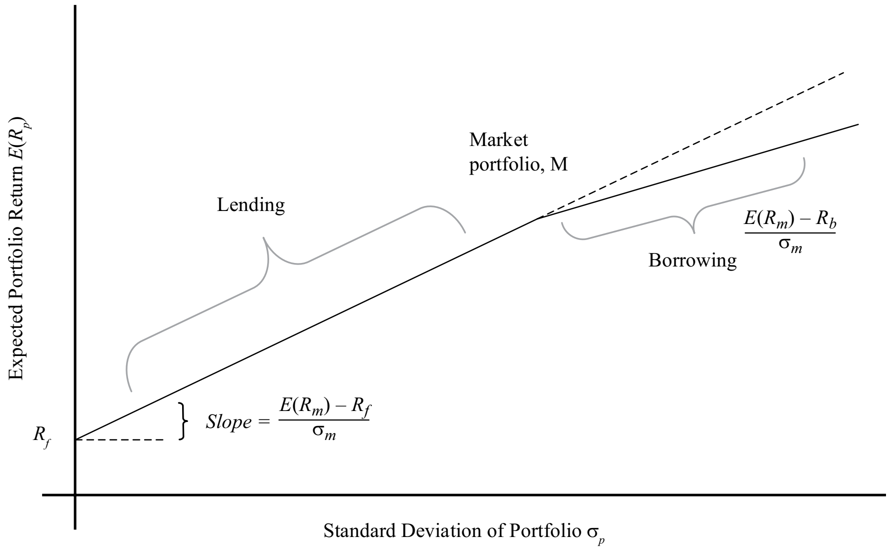
Note that negative investment (borrowing) in the risk-free asset provides a higher expected return for the portfolio but that higher return is also associated with higher risk.

4.1 Leveraged Portfolios with Different Lending and Borrowing Rates

Although we assumed that Mr. Miles can borrow at the same rate as the US government, it is more likely that he will have to pay a higher interest rate than the government because his ability to repay is not as certain as that of the government. Now consider that although Mr. Miles can invest (lend) at R_f, he can borrow at only R_b, a rate that is higher than the risk-free rate.

With different lending and borrowing rates, the CML will no longer be a single straight line. The line will have a slope of $[E(R_m) - R_f]/\sigma_m$ between Points R_f and M, where the lending rate is R_f, but will have a smaller slope of $[E(R_m) - R_b]/\sigma_m$ at points to the right of M, where the borrowing rate is R_b. Exhibit 5 illustrates the CML with different lending and borrowing rates.

Exhibit 5 CML with Different Lending and Borrowing Rates



The equations for the two lines are given below.

$$w_1 \geq 0\text{: } E(R_p) = R_f + \left(\frac{E(R_m) - R_f}{\sigma_m}\right) \times \sigma_p$$

and

$$w_1 < 0\text{: } E(R_p) = R_b + \left(\frac{E(R_m) - R_b}{\sigma_m}\right) \times \sigma_p$$

The first equation is for the line where the investment in the risk-free asset is zero or positive—that is, at M or to the left of M in Exhibit 5. The second equation is for the line where borrowing, or negative investment in the risk-free asset, occurs. Note that the only difference between the two equations is in the interest rates used for borrowing and lending.

All passive portfolios will lie on the kinked CML, although the investment in the risk-free asset may be positive (lending), zero (no lending or borrowing), or negative (borrowing). Leverage allows less risk-averse investors to increase the amount of risk they take by borrowing money and investing more than 100 percent in the passive portfolio.

EXAMPLE 3

Leveraged Portfolio with Different Lending and Borrowing Rates

Mr. Miles approaches his broker to borrow money against securities held in his portfolio. Even though Mr. Miles' loan will be secured by the securities in his portfolio, the broker's rate for lending to customers is 7 percent. Assuming a risk-free rate of 5 percent and a market return of 15 percent with a standard deviation of 20 percent, estimate Mr. Miles' expected return and risk if he invests 25 percent and 75 percent in the market and if he decides to borrow 25 percent and 75 percent of his initial investment and invest the money in the market.

Solution:

The unleveraged portfolio's standard deviation and return are calculated using the same equations as before:

$$E(R_p) = w_1 R_f + (1 - w_1)E(R_m),$$

and

$$\sigma_p = (1 - w_1)\sigma_m$$

The results are unchanged. The slope of the line for the unleveraged portfolio is 0.50, just as before:

Return with 25 percent invested in the market = (0.75 × 5%) + (0.25 × 15%) = 7.5%.

Standard deviation with 25 percent invested in the market = 0.25 × 20% = 5%.

Return with 75 percent invested in the market = (0.25 × 5%) + (0.75 × 15%) = 12.5%.

Standard deviation with 75 percent invested in the market = 0.75 × 20% = 15%.

For the leveraged portfolio, everything remains the same except that R_f is replaced with R_b.

$$E(R_p) = w_1 R_b + (1 - w_1)E(R_m),$$

and

$$\sigma_p = (1 - w_1)\sigma_{m.}$$

Return with w_1 = –0.25 = (–0.25 × 7%) + (1.25 × 15%) = 17.0%.

Standard deviation with w_1 = –0.25 = 1.25 × 20% = 25%.

Return with w_1 = –0.75 = (–0.75 × 7%) + (1.75 × 15%) = 21.0%.

Standard deviation with w_1 = –0.75 = 1.75 × 20% = 35%.

The risk and return of the leveraged portfolio is higher than that of the unleveraged portfolio. As Mr. Miles borrows more money to invest in the market, the expected return increases but so does the standard deviation of the portfolio.

The slope of the line for the leveraged portfolio is 0.40, compared with 0.50 for the unleveraged portfolio, which means that for every 1 percent increase in risk, the investor gets a 0.40 percent increase in expected return in the leveraged part of the portfolio, compared with a 0.50 percent increase in expected return in the unleveraged part of the portfolio. Only investors who are less risk averse will choose leveraged portfolios.

5 SYSTEMATIC AND NONSYSTEMATIC RISK

c explain systematic and nonsystematic risk, including why an investor should not expect to receive additional return for bearing nonsystematic risk;

In constructing a portfolio, it is important to understand the concept of correlation and how less than perfect correlation can diversify the risk of a portfolio. As a consequence, the risk of an asset held alone may be greater than the risk of that same asset when it is part of a portfolio. Because the risk of an asset varies from one environment to another, which kind of risk should an investor consider and how should that risk be priced? This section addresses the question of pricing of risk by decomposing the total risk of a security or a portfolio into systematic and nonsystematic risk. The meaning of these risks, how they are computed, and their relevance to the pricing of assets are also discussed.

5.1 Systematic Risk and Nonsystematic Risk

Systematic risk, also known as non-diversifiable or market risk, is the risk that affects the entire market or economy. In contrast, nonsystematic risk is the risk that pertains to a single company or industry and is also known as company-specific, industry-specific, diversifiable, or idiosyncratic risk.

Systematic risk is risk that cannot be avoided and is inherent in the overall market. It is non-diversifiable because it includes risk factors that are innate within the market and affect the market as a whole. Examples of factors that constitute systematic risk include interest rates, inflation, economic cycles, political uncertainty, and widespread natural disasters. These events affect the entire market, and there is no way to avoid their effect. Systematic risk can be magnified through selection or by using leverage, or diminished by including securities that have a low correlation with the portfolio, assuming they are not already part of the portfolio.

Nonsystematic risk is risk that is local or limited to a particular asset or industry that need not affect assets outside of that asset class. Examples of nonsystematic risk could include the failure of a drug trial or an airliner crash. All these events will directly affect their respective companies and possibly industries, but have no effect on assets that are far removed from these industries. Investors can avoid nonsystematic risk through diversification by forming a portfolio of assets that are not highly correlated with one another.

We will derive expressions for each kind of risk later in this reading. You will see that the sum of systematic variance and nonsystematic variance equals the total variance of the security or portfolio:

Total variance = Systematic variance + Nonsystematic variance

Although the equality relationship is between variances, you will find frequent references to total risk as the sum of systematic risk and nonsystematic risk. In those cases, the statements refer to variance, not standard deviation.

5.1.1 *Pricing of Risk*

Pricing or valuing an asset is equivalent to estimating its expected rate of return. If an asset has a known terminal value, such as the face value of a bond, then a lower current price implies a higher future return and a higher current price implies a lower future return. The relationship between price and return can also be observed in the valuation expression shown in Section 2. Therefore, we will occasionally use price and return interchangeably when discussing the price of risk.

Consider an asset with both systematic and nonsystematic risk. Assume that both kinds of risk are priced—that is, you receive a return for both systematic risk and nonsystematic risk. What will you do? Realizing that nonsystematic risk can be diversified away, you would buy assets that have a large amount of nonsystematic risk. Once you have bought those assets with nonsystematic risk, you would diversify, or reduce that risk, by including other assets that are not highly correlated. In the process, you will minimize nonsystematic risk and eventually eliminate it altogether from your portfolio. You would now have a diversified portfolio with only systematic risk, yet you would be compensated for nonsystematic risk that you no longer have. Just like everyone else, you would have an incentive to take on more and more diversifiable risk because you are compensated for it even though you can get rid of it. The demand for diversifiable risk would keep increasing until its price becomes infinite and its expected return falls to zero. This means that our initial assumption of a non-zero return for diversifiable risk was incorrect and that the correct assumption is zero return for diversifiable risk. Therefore, according to theory, in an efficient market no incremental reward is earned for taking on diversifiable risk.

We have argued that investors should not be compensated for taking on nonsystematic risk. Therefore, investors who have nonsystematic risk must diversify it away by investing in many industries, many countries, and many asset classes. Because future returns are unknown and it is not possible to pick only winners, diversification helps in offsetting poor returns in one asset class by garnering good returns in another asset class, thereby reducing the overall risk of the portfolio. In contrast, investors must be compensated for accepting systematic risk because that risk cannot be diversified away. If investors do not receive a return commensurate with the amount of systematic risk they are taking, they will refuse to accept systematic risk.

In summary, according to theory, systematic or non-diversifiable risk is priced and investors are compensated for holding assets or portfolios based only on that investment's systematic risk. Investors do not receive any return for accepting non-systematic or diversifiable risk. Therefore, it is in the interest of risk-averse investors to hold only well-diversified portfolios.

EXAMPLE 4

Systematic and Nonsystematic Risk

1. Describe the systematic and nonsystematic risk components of the following assets:
   - **A** A risk-free asset, such as a three-month Treasury bill
   - **B** The market portfolio, such as the S&P 500.
2. Consider two assets, A and B. Asset A has twice the amount of total risk as Asset B. For Asset A, systematic risk comprises two-thirds of total risk. For Asset B, all of total risk is systematic risk. Which asset should have a higher expected rate of return?

Solution to 1A:

By definition, a risk-free asset has no risk. Therefore, a risk-free asset has zero systematic risk and zero nonsystematic risk.

Solution to 1B:

As we mentioned earlier, a market portfolio is a diversified portfolio, one in which no more risk can be diversified away. We have also described it as an efficient portfolio. Therefore, a market portfolio does not contain any nonsystematic risk.

Solution to 2:

Based on the facts given, Asset A's systematic risk is one-third greater than Asset B's systematic risk. Because only systematic risk is priced or receives a return, the expected rate of return must be higher for Asset A.

6 RETURN GENERATING MODELS

d explain return generating models (including the market model) and their uses;

As previously mentioned, in order to form the market portfolio, you should combine all available risky assets. Knowledge of the correlations among those assets allows us to estimate portfolio risk. You also learned that a fully diversified portfolio will include all asset classes and essentially all assets in those asset classes. The work required for construction of the market portfolio is formidable. For example, for a portfolio of 1,000 assets, we will need 1,000 return estimates, 1,000 standard deviation estimates, and 499,500 (1,000 × 999 ÷ 2) correlations. Other related questions that arise with this analysis are whether we really need all 1,000 assets and what happens if there are errors in these estimates.

An alternate method of constructing an optimal portfolio is simpler and easier to implement. An investor begins with a known portfolio, such as the S&P 500, and then adds other assets one at a time on the basis of the asset's standard deviation, expected return, and impact on the portfolio's risk and return. This process continues until the addition of another asset does not have a significant impact on the performance of the portfolio. The process requires only estimates of systematic risk for each asset because investors will not be compensated for nonsystematic risk. Expected returns can be calculated by using return-generating models, as we will discuss in this section. In addition to using return-generating models, we will also decompose total variance into systematic variance and nonsystematic variance and establish a formal relationship between systematic risk and return. In the next section, we will expand on this discussion and introduce the CAPM as the preferred return-generating model.

6.1 Return-Generating Models

A **return-generating model** is a model that can provide an estimate of the expected return of a security given certain parameters. If systematic risk is the only relevant parameter for return, then the return-generating model will estimate the expected return for any asset given the level of systematic risk.

As with any model, the quality of estimates of expected return will depend on the quality of input estimates and the accuracy of the model. Because it is difficult to decide which factors are appropriate for generating returns, the most general form of a return-generating model is a multi-factor model. A **multi-factor model** allows more than one variable to be considered in estimating returns and can be built using different kinds of factors, such as macroeconomic, fundamental, and statistical factors.

Macroeconomic factor models use economic factors that are correlated with security returns. These factors may include economic growth, the interest rate, the inflation rate, productivity, employment, and consumer confidence. Past relationships with returns are estimated to obtain parameter estimates, which are, in turn, used for computing expected returns. Fundamental factor models analyze and use relationships between security returns and the company's underlying fundamentals, such as, for example, earnings, earnings growth, cash flow generation, investment in research, advertising, and number of patents. Finally, in a statistical factor model, historical and cross-sectional return data are analyzed to identify factors that explain variance or covariance in observed returns. These statistical factors, however, may or may not have an economic or fundamental connection to returns. For example, the conference to which the American football Super Bowl winner belongs, whether the American Football Conference or the National Football Conference, may be a factor in US stock returns, but no obvious economic connection seems to exist between the winner's conference and US stock returns. Moreover, data mining may generate many spurious factors that are devoid of any economic meaning. Because of this limitation, analysts prefer the macroeconomic and fundamental factor models for specifying and estimating return-generating models.

A general return-generating model is expressed in the following manner:

$$E(R_i) - R_f = \sum_{j=1}^{k} \beta_{ij} E(F_j) = \beta_{i1}\left[E(R_m) - R_f\right] + \sum_{j=2}^{k} \beta_{ij} E(F_j)$$

The model has k factors, $E(F_1)$, $E(F_2)$, ... $E(F_k)$. The coefficients, β_{ij}, are the factor weights (sometimes called factor loadings) associated with each factor. The left-hand side of the model has the expected excess return (i.e., the expected return over the risk-free rate). The right-hand side provides the risk factors that would generate the return or premium required to assume that risk. We have separated out one factor, $E(R_m)$, which represents the market return. All models contain return on the market portfolio as a key factor.

Three-Factor and Four-Factor Models

Eugene Fama and Kenneth French[1] suggested that a return-generating model for stock returns should include relative market capitalization of the company ("size") relative book-to-market value of the company in addition to beta. Fama and French found that past returns could be explained better with their model than with other models available at that time, most notably, the capital asset pricing model. Mark Carhart (1997) extended the Fama and French model by adding another factor: momentum, defined as relative past stock returns.

The Single-Index Model

The simplest form of a return-generating model is a single-factor linear model, in which only one factor is considered. The most common implementation is a single-index model, which uses the market factor in the following form: $E(R_i) - R_f = \beta_i[E(R_m) - R_f]$.

1 Fama and French (1992).

Although the single-index model is simple, it fits nicely with the capital market line. Recall that the CML is linear, with an intercept of R_f and a slope of $[E(R_m) - R_f]/\sigma_m$. We can rewrite the CML by moving the intercept to the left-hand side of the equation, rearranging the terms, and generalizing the subscript from p to i, for any security:

$$E(R_i) - R_f = \left(\frac{\sigma_i}{\sigma_m}\right)\left[E(R_m) - R_f\right]$$

The factor loading or factor weight, σ_i/σ_m, refers to the ratio of total security risk to total market risk. To obtain a better understanding of factor loading and to illustrate that the CML reduces to a single-index model, we decompose total risk into its components.

6.2 Decomposition of Total Risk for a Single-Index Model

With the introduction of return-generating models, particularly the single-index model, we are able to decompose total variance into systematic and nonsystematic variances. Instead of using expected returns in the single index, let us use realized returns. The difference between expected returns and realized returns is attributable to non-market changes, as an error term, e_i, in the second equation below:

$$E(R_i) - R_f = \beta_i[E(R_m) - R_f]$$

and

$$R_i - R_f = \beta_i(R_m - R_f) + e_i$$

The variance of realized returns can be expressed in the equation below (note that R_f is a constant). We can further drop the covariance term in this equation because, by definition, any non-market return is uncorrelated with the market. Thus, we are able to decompose total variance into systematic and nonsystematic variances in the second equation below:

$$\sigma_i^2 = \beta_i^2\sigma_m^2 + \sigma_e^2 + 2\text{Cov}(R_m, e_i)$$

Total variance = Systematic variance + Nonsystematic variance, which can be written as

$$\sigma_i^2 = \beta_i^2\sigma_m^2 + \sigma_e^2$$

Total risk can be expressed as

$$\sigma_i = \sqrt{\beta_i^2\sigma_m^2 + \sigma_e^2}$$

Because nonsystematic risk is zero for well-diversified portfolios, such as the market portfolio, the total risk of a market portfolio and other similar portfolios is only systematic risk, which is $\beta_i\sigma_m$. We can now return to the CML discussed in the previous subsection and replace σ_i with $\beta_i\sigma_m$ because the CML assumes that the market is a diversified portfolio. By making this substitution for the above equation, we get the following single-index model:

$$E(R_i) - R_f = \left(\frac{\sigma_i}{\sigma_m}\right) \times \left[E(R_m) - R_f\right] = \left(\frac{\beta_i\sigma_m}{\sigma_m}\right) \times \left[E(R_m) - R_f\right],$$

$$E(R_i) - R_f = \beta_i[E(R_m) - R_f]$$

Thus, the CML, which holds only for well-diversified portfolios, is fully consistent with a single-index model.

In summary, total variance may be decomposed into systematic and nonsystematic variances and the CML is the same as a single-index model for diversified portfolios.

6.3 Return-Generating Models: The Market Model

The most common implementation of a single-index model is the **market model**, in which the market return is the single factor or single index. In principle, the market model and the single-index model are similar. The difference is that the market model is easier to work with and is normally used for estimating beta risk and computing abnormal returns. The market model is

$$R_i = \alpha_i + \beta_i R_m + e_i$$

To be consistent with the previous section, $\alpha_i = R_f(1 - \beta)$. The intercept, α_i, and slope coefficient, β_i, can be estimated by using historical security and market returns. These parameter estimates are then used to predict company-specific returns that a security may earn in a future period. Assume that a regression of Wal-Mart's historical daily returns on S&P 500 daily returns gives an α_i of 0.0001 and a β_i of 0.9. Thus, Wal-Mart's expected daily return = 0.0001 + 0.90 × R_m. If, on a given day the market rises by 1 percent and Wal-Mart's stock rises by 2 percent, then Wal-Mart's company-specific return (e_i) for that day = $R_i - E(R_i) = R_i - (\alpha_i + \beta_i R_m)$ = 0.02 – (0.0001 + 0.90 × 0.01) = 0.0109, or 1.09%. In other words, Wal-Mart earned an abnormal return of 1.09 percent on that day.

7 CALCULATION AND INTERPRETATION OF BETA

e calculate and interpret beta;

We begin with the single-index model introduced earlier using realized returns and rewrite it as

$$R_i = (1 - \beta_i)R_f + \beta_i \times R_m + e_i$$

Because systematic risk depends on the correlation between the asset and the market, we can arrive at a measure of systematic risk from the covariance between R_i and R_m, where R_i is defined using the above equation. Note that the risk-free rate is a constant, so the first term in R_i drops out.

$$\begin{aligned}\mathrm{Cov}(R_i, R_m) &= \mathrm{Cov}(\beta_i \times R_m + e_i, R_m)\\ &= \beta_i \mathrm{Cov}(R_m, R_m) + \mathrm{Cov}(e_i, R_m)\\ &= \beta_i \sigma_m^2 + 0\end{aligned}$$

The first term is beta multiplied by the variance of R_m. Because the error term is uncorrelated with the market, the second term drops out. Then, we can rewrite the equation in terms of beta as follows:

$$\beta_i = \frac{\mathrm{Cov}(R_i, R_m)}{\sigma_m^2} = \frac{\rho_{i,m}\sigma_i\sigma_m}{\sigma_m^2} = \frac{\rho_{i,m}\sigma_i}{\sigma_m}$$

The above formula shows the expression for beta, β_i, which is similar to the factor loading in the single-index model presented earlier. For example, if the correlation between an asset and the market is 0.70 and the asset and market have standard deviations of return of 0.25 and 0.15, respectively, the asset's beta would be (0.70)(0.25)/0.15 = 1.17. If the asset's covariance with the market and market variance were given as 0.026250 and 0.02250, respectively, the calculation would be 0.026250/0.02250 = 1.17. The beta in the market model includes an adjustment for the correlation between asset i and the market because the market model covers all assets whereas the CML works only for fully diversified portfolios.

As shown in the above equation, **beta** is a measure of how sensitive an asset's return is to the market as a whole and is calculated as the covariance of the return on i and the return on the market divided by the variance of the market return; that expression is equivalent to the product of the asset's correlation with the market with a ratio of standard deviations of return (i.e., the ratio of the asset's standard deviation to the market's). As we have shown, beta captures an asset's systematic risk, or the portion of an asset's risk that cannot be eliminated by diversification. The variances and correlations required for the calculation of beta are usually based on historical returns.

A positive beta indicates that the return of an asset follows the general market trend, whereas a negative beta shows that the return of an asset generally follows a trend that is opposite to that of the market. In other words, a positive beta indicates that the return of an asset moves in the same direction of the market, whereas a negative beta indicates that the return of an asset moves in the opposite direction of the market. A risk-free asset's beta is zero because its covariance with other assets is zero. In other words, a beta of zero indicates that the asset's return has no correlation with movements in the market. The market's beta can be calculated by substituting σ_m for σ_i in the numerator. Also, any asset's correlation with itself is 1, so the beta of the market is 1:

$$\beta_i = \frac{\rho_{i,m}\sigma_i}{\sigma_m} = \frac{\rho_{m,m}\sigma_m}{\sigma_m} = 1$$

Because the market's beta is 1, the average beta of stocks in the market, by definition, is 1. In terms of correlation, most stocks, especially in developed markets, tend to be highly correlated with the market, with correlations in excess of 0.70. Some US broad market indexes, such as the S&P 500, the Dow Jones 30, and the NASDAQ 100, have even higher correlations that are in excess of 0.90. The correlations among different sectors are also high, which shows that companies have similar reactions to the same economic and market changes. As a consequence and as a practical matter, finding assets that have a consistently negative beta is unusual because of the market's broad effects on all assets.

EXAMPLE 5

Calculation of Beta

Assuming that the risk (standard deviation) of the market is 25 percent, calculate the beta for the following assets:

1 A short-term US Treasury bill.
2 Gold, which has a standard deviation equal to the standard deviation of the market but a zero correlation with the market.
3 A new emerging market that is not currently included in the definition of "market"—the emerging market's standard deviation is 60 percent, and the correlation with the market is –0.1.
4 An initial public offering or new issue of stock with a standard deviation of 40 percent and a correlation with the market of 0.7 (IPOs are usually very risky but have a relatively low correlation with the market).

We use the formula for beta in answering the above questions: $\beta_i = \dfrac{\rho_{i,m}\sigma_i}{\sigma_m}$

Solution to 1:

By definition, a short-term US Treasury bill has zero risk. Therefore, its beta is zero.

Solution to 2:

Because the correlation of gold with the market is zero, its beta is zero.

Solution to 3:

Beta of the emerging market is $-0.1 \times 0.60 \div 0.25 = -0.24$.

Solution to 4:

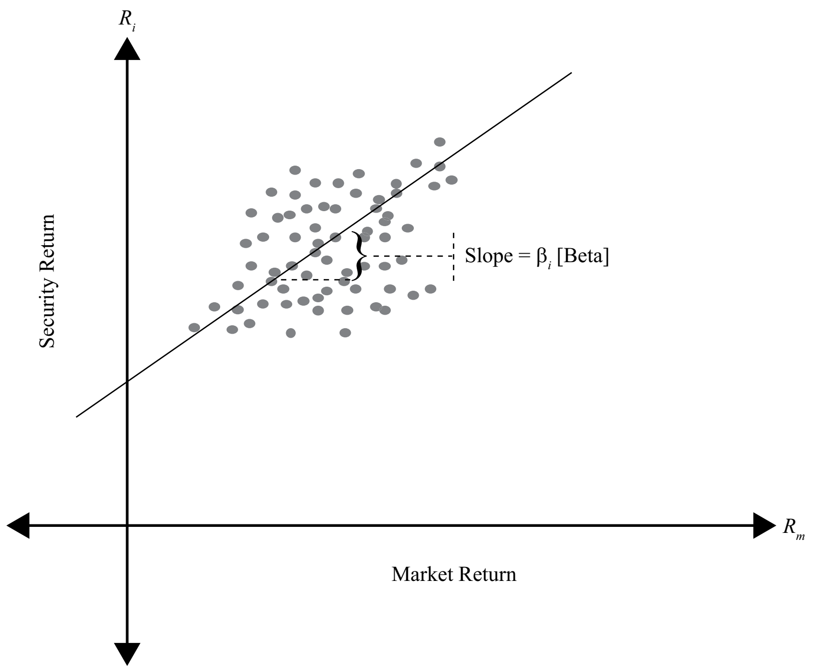
Beta of the initial public offering is $0.7 \times 0.40 \div 0.25 = 1.12$.

7.1 Estimation of Beta

An alternative and more practical approach is to estimate beta directly by using the market model described above. The market model, $R_i = \alpha_i + \beta_i R_m + e_i$, is estimated by using regression analysis, which is a statistical process that evaluates the relationship between a given variable (the dependent variable) and one or more other (independent) variables. Historical security returns (R_i) and historical market returns (R_m) are inputs used for estimating the two parameters α_i and β_i.

Regression analysis is similar to plotting all combinations of the asset's return and the market return (R_i, R_m) and then drawing a line through all points such that it minimizes the sum of squared linear deviations from the line. Exhibit 6 illustrates the market model and the estimated parameters. The intercept, α_i (sometimes referred to as the constant), and the slope term, β_i, are all that is needed to define the security characteristic line and obtain beta estimates.

Exhibit 6 Beta Estimation Using a Plot of Security and Market Returns



Although beta estimates are important for forecasting future levels of risk, there is much concern about their accuracy. In general, shorter periods of estimation (e.g., 12 months) represent betas that are closer to the asset's current level of systematic risk. Shorter period beta estimates, however, are also less accurate than beta estimates measured over three to five years because they may be affected by special events in that short period. Although longer period beta estimates are more accurate, they may be a poor representation of future expectations, especially if major changes in the asset have occurred. Therefore, it is necessary to recognize that estimates of beta, whether obtained through calculation or regression analysis, may or may not represent current or future levels of an asset's systematic risk.

7.2 Beta and Expected Return

Although the single-index model, also called the **capital asset pricing model** (CAPM), will be discussed in greater detail in the next section, we will use the CAPM in this section to estimate returns, given asset betas. The CAPM is usually written with the risk-free rate on the right-hand side:

$$E(R_i) = R_f + \beta_i[E(R_m) - R_f]$$

The model shows that the primary determinant of expected return for a security is its beta, or how well the security correlates with the market. The higher the beta of an asset, the higher its expected return will be. Assets with a beta greater than 1 have an expected return that is higher than the market return, whereas assets with a beta of less than 1 have an expected return that is less than the market return.

In certain cases, assets may require a return less than the risk-free return. For example, if an asset's beta is negative, the required return will be less than the risk-free rate. When combined with the market, the asset reduces the risk of the overall portfolio, which makes the asset very valuable. Insurance is one such asset. Insurance gives a positive return when the insured's wealth is reduced because of a catastrophic loss. In the absence of such a loss or when the insured's wealth is growing, the insured is required to pay an insurance premium. Thus, insurance has a negative beta and a negative expected return, but helps in reducing overall risk.

EXAMPLE 6

Calculation of Expected Return

1 Alpha Natural Resources (ANR), a coal producer, buys a large but privately held coal producer in China. As a result of the cross-border acquisition of a private company, ANR's standard deviation of returns is reduced from 50 percent to 30 percent and its correlation with the market falls from 0.95 to 0.75. Assume that the standard deviation and return of the market remain unchanged at 25 percent and 10 percent, respectively, and that the risk-free rate is 3 percent.

A Calculate the beta of ANR stock and its expected return before the acquisition.

B Calculate the expected return after the acquisition.

Solution to 1A:

Using the formula for β_i, we can calculate β_i and then the return.

$$\beta_i = \frac{\rho_{i,m}\sigma_i}{\sigma_m} = \frac{0.95 \times 0.50}{0.25} = 1.90$$

$$E(R_i) = R_f + \beta_i[E(R_m) - R_f] = 0.03 + 1.90 \times (0.10 - 0.03) = 0.163 = 16.3\%$$

Solution to 1B:

We follow the same procedure but with the after-acquisition correlation and risk.

$$\beta_i = \frac{\rho_{i,m}\sigma_i}{\sigma_m} = \frac{0.75 \times 0.30}{0.25} = 0.90$$

$$E(R_i) = R_f + \beta_i[E(R_m) - R_f] = 0.03 + 0.90 \times (0.10 - 0.03) = 0.093 = 9.3\%$$

The market risk premium is 7 percent (10% – 3%). As the beta changes, the change in the security's expected return is the market risk premium multiplied by the change in beta. In this scenario, ANR's beta decreased by 1.0, so the new expected return for ANR is 7 percentage points lower.

2 Mr. Miles observes the strong demand for iPods and iPhones and wants to invest in Apple stock. Unfortunately, Mr. Miles doesn't know the return he should expect from his investment. He has been given a risk-free rate of 3 percent, a market return of 10 percent, and Apple's beta of 1.5.

- **A** Calculate Apple's expected return.
- **B** An analyst looking at the same information decides that the past performance of Apple is not representative of its future performance. He decides that, given the increase in Apple's market capitalization, Apple acts much more like the market than before and thinks Apple's beta should be closer to 1.1. What is the analyst's expected return for Apple stock?

Solution to 2A:

$$E(R_i) = R_f + \beta_i[E(R_m) - R_f] = 0.03 + 1.5 \times (0.10 - 0.03) = 0.135 = 13.5\%$$

Solution to 2B:

$$E(R_i) = R_f + \beta_i[E(R_m) - R_f] = 0.03 + 1.1 \times (0.10 - 0.03) = 0.107 = 10.7\%$$

This example illustrates the lack of connection between estimation of past returns and projection into the future. Investors should be aware of the limitations of using past returns for estimating future returns.

CAPITAL ASSET PRICING MODEL: ASSUMPTIONS AND THE SECURITY MARKET LINE

8

- **f** explain the capital asset pricing model (CAPM), including its assumptions, and the security market line (SML);
- **g** calculate and interpret the expected return of an asset using the CAPM;

The capital asset pricing model is one of the most significant innovations in portfolio theory. The model is simple, yet powerful; is intuitive, yet profound. The CAPM was introduced independently by William Sharpe, John Lintner, Jack Treynor, and Jan Mossin and builds on Harry Markowitz's earlier work on diversification and modern portfolio theory.[2] The model provides a linear expected return–beta relationship that

2 See, for example, Markowitz (1952), Sharpe (1964), Lintner (1965a, 1965b), Treynor (1961, 1962), and Mossin (1966).

precisely determines the expected return given the beta of an asset. In doing so, it makes the transition from total risk to systematic risk, the primary determinant of expected return. Recall the following equation:

$$E(R_i) = R_f + \beta_i[E(R_m) - R_f]$$

The CAPM asserts that the expected returns of assets vary only by their systematic risk as measured by beta. Two assets with the same beta will have the same expected return irrespective of the nature of those assets. Given the relationship between risk and return, all assets are defined only by their beta risk, which we will explain as the assumptions are described.

In the remainder of this section, we will examine the assumptions made in arriving at the CAPM and the limitations those assumptions entail. Second, we will implement the CAPM through the security market line to price any portfolio or asset, both efficient and inefficient. Finally, we will discuss ways in which the CAPM can be applied to investments, valuation, and capital budgeting.

8.1 Assumptions of the CAPM

Similar to all other models, the CAPM ignores many of the complexities of financial markets by making simplifying assumptions. These assumptions allow us to gain important insights into how assets are priced without complicating the analysis. Once the basic relationships are established, we can relax the assumptions and examine how our insights need to be altered. Some of these assumptions are constraining, whereas others are benign. And other assumptions affect only a particular set of assets or only marginally affect the hypothesized relationships.

1 ***Investors are risk-averse, utility-maximizing, rational individuals.***

Risk aversion means that investors expect to be compensated for accepting risk. Note that the assumption does not require investors to have the same degree of risk aversion; it only requires that they are averse to risk. Utility maximization implies that investors want higher returns, not lower returns, and that investors always want more wealth (i.e., investors are never satisfied). Investors are understood to be rational in that they correctly evaluate and analyze available information to arrive at rational decisions. Although rational investors may use the same information to arrive at different estimates of expected risk and expected returns, homogeneity among investors (see Assumption 4) requires that investors be rational individuals.

Risk aversion and utility maximization are generally accepted as reflecting a realistic view of the world. Yet, rationality among investors has been questioned because investors may allow their personal biases and experiences to disrupt their decision making, resulting in suboptimal investments. Nonetheless, the model's results are unaffected by such irrational behavior as long as it does not affect prices in a significant manner (i.e., the trades of irrational investors cancel each other or are dominated by the trades of rational investors).

2 ***Markets are frictionless, including no transaction costs and no taxes.***

Frictionless markets allow us to abstract the analysis from the operational characteristics of markets. In doing so, we do not allow the risk–return relationship to be affected by, for example, the trading volume on the New York Stock Exchange or the difference between buying and selling prices. Specifically, frictionless markets do not have transaction costs, taxes, or any costs or restrictions on short selling. We also assume that borrowing and lending at the risk-free rate is possible.

The transaction costs of many large institutions are negligible, and many institutions do not pay taxes. Even the presence of non-zero transaction costs, taxes, or the inability to borrow at the risk-free rate does not materially affect the general conclusions of the CAPM. Costs of short selling or restrictions on short selling, however, can introduce an upward bias in asset prices, potentially jeopardizing important conclusions of the CAPM.

3 ***Investors plan for the same single holding period.***

The CAPM is a single-period model, and all investor decisions are made on the basis of that one period. The assumption of a single period is applied for convenience because working with multi-period models is more difficult. A single-period model, however, does not allow learning to occur, and bad decisions can persist. In addition, maximizing utility at the end of a multi-period horizon may require decisions in certain periods that may seem suboptimal when examined from a single-period perspective. Nonetheless, the single holding period does not severely limit the applicability of the CAPM to multi-period settings.

4 ***Investors have homogeneous expectations or beliefs.***

This assumption means that all investors analyze securities in the same way using the same probability distributions and the same inputs for future cash flows. In addition, given that they are rational individuals, the investors will arrive at the same valuations. Because their valuations of all assets are identical, they will generate the same optimal risky portfolio, which we call the market portfolio.

The assumption of homogeneous beliefs can be relaxed as long as the differences in expectations do not generate significantly different optimal risky portfolios.

5 ***All investments are infinitely divisible.***

This assumption implies that an individual can invest as little or as much as he or she wishes in an asset. This supposition allows the model to rely on continuous functions rather than on discrete jump functions. The assumption is made for convenience only and has an inconsequential impact on the conclusions of the model.

6 ***Investors are price takers.***

The CAPM assumes that there are many investors and that no investor is large enough to influence prices. Thus, investors are price takers, and we assume that security prices are unaffected by investor trades. This assumption is generally true because even though investors may be able to affect prices of small stocks, those stocks are not large enough to affect the primary results of the CAPM.

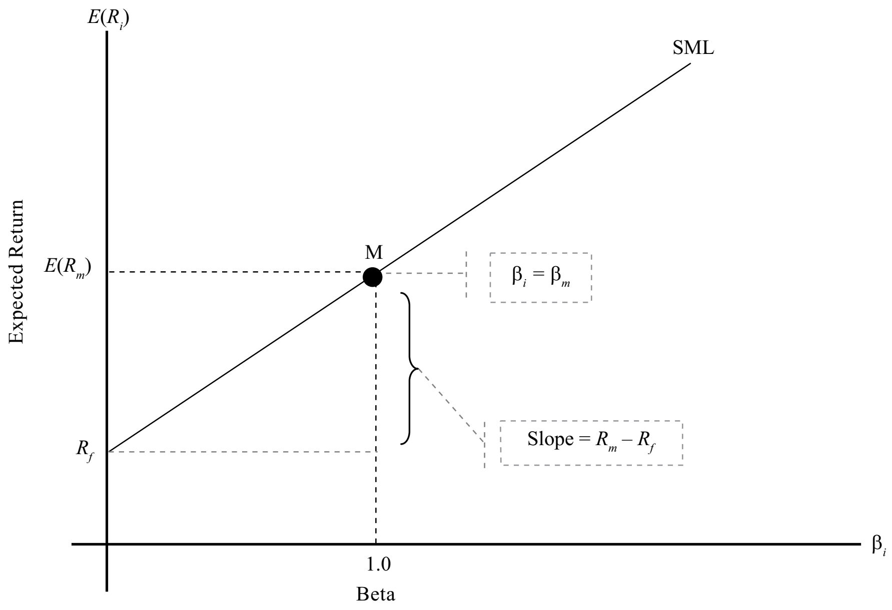
The main objective of these assumptions is to create a marginal investor who rationally chooses a mean–variance-efficient portfolio in a predictable fashion. We assume away any inefficiency in the market from both operational and informational perspectives. Although some of these assumptions may seem unrealistic, relaxing most of them will have only a minor influence on the model and its results. Moreover, the CAPM, with all its limitations and weaknesses, provides a benchmark for comparison and for generating initial return estimates.

8.2 The Security Market Line

In this subsection, we apply the CAPM to the pricing of securities. The **security market line** (SML) is a graphical representation of the capital asset pricing model with beta, reflecting systematic risk, on the x-axis and expected return on the y-axis. Using the same concept as the capital market line, the SML intersects the *y*-axis at the risk-free rate of return, and the slope of this line is the market risk premium, $R_m - R_f$. Recall that the capital market line (CML) does not apply to all securities or assets but only to portfolios on the efficient frontier. The efficient frontier gives optimal combinations of expected return and total risk. In contrast, the security market line applies to any security, efficient or not. Total risk and systematic risk are equal only for efficient portfolios because those portfolios have no diversifiable risk remaining.

Exhibit 7 is a graphical representation of the CAPM, the security market line. As shown earlier in this reading, the beta of the market is 1 (*x*-axis) and the market earns an expected return of R_m (*y*-axis). Using this line, it is possible to calculate the expected return of an asset. The next example illustrates the beta and return calculations.

Exhibit 7 The Security Market Line



EXAMPLE 7

Security Market Line and Expected Return

1 Suppose the risk-free rate is 3 percent, the expected return on the market portfolio is 13 percent, and its standard deviation is 23 percent. An Indian company, Bajaj Auto, has a standard deviation of 50 percent but is uncorrelated with the market. Calculate Bajaj Auto's beta and expected return.

2 Suppose the risk-free rate is 3 percent, the expected return on the market portfolio is 13 percent, and its standard deviation is 23 percent. A German company, Mueller Metals, has a standard deviation of 50 percent and a correlation of 0.65 with the market. Calculate Mueller Metal's beta and expected return.

Solution to 1:

Using the formula for β_i, we can calculate β_i and then the return.

$$\beta_i = \frac{\rho_{i,m}\sigma_i}{\sigma_m} = \frac{0.0 \times 0.50}{0.23} = 0$$

$$E(R_i) = R_f + \beta_i[E(R_m) - R_f] = 0.03 + 0 \times (0.13 - 0.03) = 0.03 = 3.0\%$$

Because of its zero correlation with the market portfolio, Bajaj Auto's beta is zero. Because the beta is zero, the expected return for Bajaj Auto is the risk-free rate, which is 3 percent.

Solution to 2:

Using the formula for β_i, we can calculate β_i and then the return.

$$\beta_i = \frac{\rho_{i,m}\sigma_i}{\sigma_m} = \frac{0.65 \times 0.50}{0.23} = 1.41$$

$$E(R_i) = R_f + \beta_i[E(R_m) - R_f] = 0.03 + 1.41\times (0.13 - 0.03) = 0.171 = 17.1\%$$

Because of the high degree of correlation with the market, the beta for Mueller Metals is 1.41 and the expected return is 17.1 percent. Because Mueller Metals has systematic risk that is greater than that of the market, it has an expected return that exceeds the expected return of the market.

8.2.1 *Portfolio Beta*

As we stated above, the security market line applies to all securities. But what about a combination of securities, such as a portfolio? Consider two securities, 1 and 2, with a weight of w_i in Security 1 and the balance in Security 2. The return for the two securities and return of the portfolio can be written as:

$$\begin{aligned}
E(R_1) &= R_f + \beta_1\left[E(R_m) - R_f\right] \\
E(R_2) &= R_f + \beta_2\left[E(R_m) - R_f\right] \\
E(R_p) &= w_1 E(R_1) + w_2 E(R_2) \\
&= w_1 R_f + w_1\beta_1\left[E(R_m) - R_f\right] + w_2 R_f + w_2\beta_2\left[E(R_m) - R_f\right] \\
&= R_f + (w_1\beta_1 + w_2\beta_2)\left[E(R_m) - R_f\right]
\end{aligned}$$

The last equation gives the expression for the portfolio's expected return. From this equation, we can conclude that the portfolio's beta = $w_1\beta_1 + w_2\beta_2$. In general, the portfolio beta is a weighted sum of the betas of the component securities and is given by:

$$\beta_p = \sum_{i=1}^{n} w_i\beta_i;\ \sum_{i=1}^{n} w_i = 1$$

The portfolio's return given by the CAPM is

$$E(R_p) = R_f + \beta_p[E(R_m) - R_f]$$

This equation shows that a linear relationship exists between the expected return of a portfolio and the systematic risk of the portfolio as measured by β_p.

EXAMPLE 8

Portfolio Beta and Return

You invest 20 percent of your money in the risk-free asset, 30 percent in the market portfolio, and 50 percent in RedHat, a US stock that has a beta of 2.0. Given that the risk-free rate is 4 percent and the market return is 16 percent, what are the portfolio's beta and expected return?

Solution:

The beta of the risk-free asset = 0, the beta of the market = 1, and the beta of RedHat is 2.0. The portfolio beta is

$$\beta_p = w_1\beta_1 + w_2\beta_2 + w_3\beta_3 = (0.20 \times 0.0) + (0.30 \times 1.0) + (0.50 \times 2.0) = 1.30$$

$$E(R_i) = R_f + \beta_i[E(R_m) - R_f] = 0.04 + 1.30 \times (0.16 - 0.04) = 0.196 = 19.6\%$$

The portfolio beta is 1.30, and its expected return is 19.6 percent.

Alternate Method:

Another method for calculating the portfolio's return is to calculate individual security returns and then use the portfolio return formula (i.e., weighted average of security returns) to calculate the overall portfolio return.

Return of the risk-free asset = 4 percent; return of the market = 16 percent

RedHat's return based on its beta = 0.04 + 2.0 × (0.16 – 0.04) = 0.28

Portfolio return = (0.20 × 0.04) + (0.30 × 0.16) + (0.50 × 0.28) = 0.196 = 19.6%

Not surprisingly, the portfolio return is 19.6 percent, as calculated in the first method.

9 CAPITAL ASSET PRICING MODEL: APPLICATIONS

g calculate and interpret the expected return of an asset using the CAPM;

h describe and demonstrate applications of the CAPM and the SML;

The CAPM offers powerful and intuitively appealing predictions about risk and the relationship between risk and return. The CAPM is not only important from a theoretical perspective but is also used extensively in practice. In this section, we will discuss some common applications of the model. When applying these tools to different scenarios, it is important to understand that the CAPM and the SML are functions that give an indication of what the return in the market *should* be, given a certain level of risk. The actual return may be quite different from the expected return.

Applications of the CAPM include estimates of the expected return for capital budgeting, comparison of the actual return of a portfolio or portfolio manager with the CAPM return for performance appraisal, and the analysis of alternate return estimates and the CAPM returns as the basis for security selection. The applications are discussed in more detail in this section.

9.1 Estimate of Expected Return

Given an asset's systematic risk, the expected return can be calculated using the CAPM. Recall that the price of an asset is the sum of all future cash flows discounted at the required rate of return, where the discount rate or the required rate of return is commensurate with the asset's risk. The expected rate of return obtained from the CAPM is normally the first estimate that investors use for valuing assets, such as stocks, bonds, real estate, and other similar assets. The required rate of return from the CAPM is also used for capital budgeting and determining the economic feasibility of projects. Again, recall that when computing the net present value of a project, investments and net revenues are considered cash flows and are discounted at the required rate of return. The required rate of return, based on the project's risk, is calculated using the CAPM.

Because risk and return underlie almost all aspects of investment decision making, it is not surprising that the CAPM is used for estimating expected return in many scenarios. Other examples include calculating the cost of capital for regulated companies by regulatory commissions and setting fair insurance premiums. The next example shows an application of the CAPM to capital budgeting.

EXAMPLE 9

Application of the CAPM to Capital Budgeting

GlaxoSmithKline Plc is examining the economic feasibility of developing a new medicine. The initial investment in Year 1 is $500 million. The investment in Year 2 is $200 million. There is a 50 percent chance that the medicine will be developed and will be successful. If that happens, GlaxoSmithKline must spend another $100 million in Year 3, but its income from the project in Year 3 will be $500 million, not including the third-year investment. In Years 4, 5, and 6, it will earn $400 million a year if the medicine is successful. At the end of Year 6, it intends to sell all rights to the medicine for $600 million. If the medicine is unsuccessful, none of GlaxoSmithKline's investments can be salvaged. Assume that the market return is 12 percent, the risk-free rate is 2 percent, and the beta risk of the project is 2.3. All cash flows occur at the end of each year.

1 Calculate the expected annual cash flows using the probability of success.
2 Calculate the expected return.
3 Calculate the net present value.

Solution to 1:

There is a 50 percent chance that the cash flows in Years 3–6 will occur. Taking that into account, the expected annual cash flows are:

Year 1: –$500 million (outflow)

Year 2: –$200 million (outflow)

Year 3: 50% of –$100 million (outflow) + 50% of $500 million = $200 million

Year 4: 50% of $400 million = $200 million

Year 5: 50% of $400 million = $200 million

Year 6: 50% of $400 million + 50% of $600 million = $500 million

Solution to 2:

The expected or required return for the project can be calculated using the CAPM, which is $= 0.02 + 2.3 \times (0.12 - 0.02) = 0.25$.

Solution to 3:

The net present value is the discounted value of all cash flows:

$$NPV = \sum_{t=0}^{T} \frac{CF_t}{(1 + r_t)^t}$$
$$= \frac{-500}{(1 + 0.25)} + \frac{-200}{(1 + 0.25)^2} + \frac{200}{(1 + 0.25)^3} + \frac{200}{(1 + 0.25)^4}$$
$$+ \frac{200}{(1 + 0.25)^5} + \frac{500}{(1 + 0.25)^6}$$
$$= -400 - 128 + 102.40 + 81.92 + 65.54 + 131.07 = -147.07.$$

Because the net present value is negative (–$147.07 million), the project should not be accepted by GlaxoSmithKline.

10 BEYOND CAPM: LIMITATIONS AND EXTENSIONS OF CAPM

h describe and demonstrate applications of the CAPM and the SML;

In general, return-generating models allow us to estimate an asset's return given its characteristics, where the asset characteristics required for estimating the return are specified in the model. Estimating an asset's return is important for investment decision making. These models are also important as a benchmark for evaluating portfolio, security, or manager performance. The return-generating models were briefly introduced in Section 6, and one of those models, the capital asset pricing model, was discussed in detail in Sections 8 and 9.

The purpose of this section is to make readers aware that, although the CAPM is an important concept and model, the CAPM is not the only return-generating model. In this section, we revisit and highlight the limitations of the CAPM and preview return-generating models that address some of those limitations.

10.1 Limitations of the CAPM

The CAPM is subject to theoretical and practical limitations. Theoretical limitations are inherent in the structure of the model, whereas practical limitations are those that arise in implementing the model.

10.1.1 *Theoretical Limitations of the CAPM*

- Single-factor model: Only systematic risk or beta risk is priced in the CAPM. Thus, the CAPM states that no other investment characteristics should be considered in estimating returns. As a consequence, it is prescriptive and easy to understand and apply, although it is very restrictive and inflexible.
- Single-period model: The CAPM is a single-period model that does not consider multi-period implications or investment objectives of future periods, which can lead to myopic and suboptimal investment decisions. For example, it may be optimal to default on interest payments in the current period to

maximize current returns, but the consequences may be negative in the next period. A single-period model like the CAPM is unable to capture factors that vary over time and span several periods.

10.1.2 *Practical Limitations of the CAPM*

In addition to the theoretical limitations, implementation of the CAPM raises several practical concerns, some of which are listed below.

- Market portfolio: The true market portfolio according to the CAPM includes all assets, financial and nonfinancial, which means that it also includes many assets that are not investable, such as human capital and assets in closed economies. Richard Roll[3] noted that one reason the CAPM is not testable is that the true market portfolio is unobservable.
- Proxy for a market portfolio: In the absence of a true market portfolio, market participants generally use proxies. These proxies, however, vary among analysts, the country of the investor, etc. and generate different return estimates for the same asset, which is impermissible in the CAPM.
- Estimation of beta risk: A long history of returns (three to five years) is required to estimate beta risk. The historical state of the company, however, may not be an accurate representation of the current or future state of the company. More generally, the CAPM is an *ex ante* model, yet it is usually applied using *ex post* data. In addition, using different periods for estimation results in different estimates of beta. For example, a three-year beta is unlikely to be the same as a five-year beta, and a beta estimated with daily returns is unlikely to be the same as the beta estimated with monthly returns. Thus, we are likely to estimate different returns for the same asset depending on the estimate of beta risk used in the model.
- The CAPM is a poor predictor of returns: If the CAPM is a good model, its estimate of asset returns should be closely associated with realized returns. However, empirical support for the CAPM is weak.[4] In other words, tests of the CAPM show that asset returns are not determined only by systematic risk. Poor predictability of returns when using the CAPM is a serious limitation because return-generating models are used to estimate future returns.
- Homogeneity in investor expectations: The CAPM assumes that homogeneity exists in investor expectations for the model to generate a single optimal risky portfolio (the market) and a single security market line. Without this assumption, there will be numerous optimal risky portfolios and numerous security market lines. Clearly, investors can process the same information in a rational manner and arrive at different optimal risky portfolios.

10.2 Extensions to the CAPM

Given the limitations of the CAPM, it is not surprising that other models have been proposed to address some of these limitations. These new models are not without limitations of their own, which we will mention while discussing the models. We divide the models into two categories—theoretical models and practical models—and provide one example of each type.

---

3 Roll (1977).
4 See, for example, Fama and French (1992).

10.2.1 *Theoretical Models*

Theoretical models are based on the same principle as the CAPM but expand the number of risk factors. The best example of a theoretical model is the arbitrage pricing theory (APT), which was developed by Stephen Ross.[5] Like the CAPM, APT proposes a linear relationship between expected return and risk:

$$E(R_p) = R_F + \lambda_1\beta_{p,1} + \ldots + \lambda_K\beta_{p,K}$$

where

$E(R_p)$ = the expected return of portfolio p
R_F = the risk-free rate
λ_j = the risk premium (expected return in excess of the risk-free rate) for factor j
$\beta_{p,j}$ = the sensitivity of the portfolio to factor j
K = the number of risk factors

Unlike the CAPM, however, APT allows numerous risk factors—as many as are relevant to a particular asset. Moreover, other than the risk-free rate, the risk factors need not be common and may vary from one asset to another. A no-arbitrage condition in asset markets is used to determine the risk factors and estimate betas for the risk factors.

Although it is theoretically elegant, flexible, and superior to the CAPM, APT is not commonly used in practice because it does not specify any of the risk factors and it becomes difficult to identify risk factors and estimate betas for each asset in a portfolio. So from a practical standpoint, the CAPM is preferred to APT.

10.2.2 *Practical Models*

If beta risk in the CAPM does not explain returns, which factors do? Practical models seek to answer this question through extensive research. As mentioned in Section 6, the best example of such a model is the four-factor model proposed by Fama and French (1992) and Carhart (1997).

Based on an analysis of the relationship between past returns and a variety of different factors, Fama and French (1992) proposed that three factors seem to explain asset returns better than just systematic risk. Those three factors are relative size, relative book-to-market value, and beta of the asset. With Carhart's (1997) addition of relative past stock returns, the model can be written as follows:

$$E(R_{it}) = \alpha_i + \beta_{i,MKT}MKT_t + \beta_{i,SMB}SMB_t + \beta_{i,HML}HML_t + \beta_{i,UMD}UMD_t$$

where

$E(R_i)$ = the return on an asset in excess of the one-month T-bill return
MKT = the excess return on the market portfolio
SMB = the difference in returns between small-capitalization stocks and large-capitalization stocks (size)
HML = the difference in returns between high-book-to-market stocks and low-book-to-market stocks (value versus growth)
UMD = the difference in returns of the prior year's winners and losers (momentum)

5 Ross (1976).

Historical analysis shows that the coefficient on *MKT* is not significantly different from zero, which implies that stock return is unrelated to the market. The factors that explain stock returns are size (smaller companies outperform larger companies), book-to-market ratio (value companies outperform glamour companies), and momentum (past winners outperform past losers).

The four-factor model has been found to predict asset returns much better than the CAPM and is extensively used in estimating returns for US stocks.

Two observations are in order. First, the model is not underpinned by a theory of market equilibrium, as is the case for the CAPM. Second, there is no assurance that the model will continue to work well in the future.

PORTFOLIO PERFORMANCE APPRAISAL MEASURES

11

i. calculate and interpret the Sharpe ratio, Treynor ratio, M^2, and Jensen's alpha.

In the investment industry, **performance evaluation** refers to the measurement, attribution, and appraisal of investment results. In particular, performance evaluation provides information about the return and risk of investment portfolios over specified investment period(s). By providing accurate data and analysis on investment decisions and their consequences, performance evaluation allows portfolio managers to take corrective measures to improve investment decision-making and management processes. Performance evaluation information helps in understanding and controlling investment risk and should, therefore, lead to improved risk management. Performance evaluation seeks to answer the following questions:

- What was the investment portfolio's past performance, and what may be expected in the future?

Answering this question is the subject of performance measurement. *Performance measurement is concerned with the measurement of return and risk.*

- How did the investment portfolio produce its observed performance, and what are the expected sources of expected future performance?

Answering this question is the subject of performance attribution. *Performance attribution is concerned with identifying and quantifying the sources of performance of a portfolio.*

- Was the observed investment portfolio's performance the result of investment skill or luck?

Answering this question is the subject of performance appraisal. *Performance appraisal is concerned with identifying and measuring investment skill.*

The information provided by performance evaluation is of great interest to all stakeholders in the investment management process because of its value in evaluating the overall quality of the investment management process as well as individual investment decisions.

In this reading, performance appraisal is based only on the CAPM. However, it is easy to extend this analysis to multi-factor models that may include industry or other special factors. Four ratios are commonly used in performance appraisal.

11.1 The Sharpe Ratio

Performance has two components, risk and return. Although return maximization is a laudable objective, comparing just the return of a portfolio with that of the market is not sufficient. Because investors are risk averse, they will require compensation for higher risk in the form of higher returns. A commonly used measure of performance is the **Sharpe ratio**, which is defined as the portfolio's risk premium divided by its risk. An appealing feature of the Sharpe ratio is that its use can be justified on a theoretical *ex ante* (before the fact) basis and *ex post* (after the fact) values can easily be determined by using readily available market data. The Sharpe ratio is also easy to interpret, essentially being an efficiency ratio relating reward to risks taken. It is the most widely recognized and used appraisal measure.

The equation below defines the *ex ante* Sharpe ratio in terms of three inputs: (1) the portfolio's expected return, $E(R_p)$; (2) the risk-free rate of interest, R_f; and (3) the portfolio's *ex ante* standard deviation of returns (return volatility), σ_p, a quantitative measure of total risk.

$$\text{SR} = \frac{E\left(R_p\right) - R_F}{\sigma_p}$$

The Sharpe ratio can also be used on an *ex post* basis to evaluate historical risk-adjusted returns. Assume we have a sample of historical data that can be used to determine the sample mean portfolio return, $\overline{R}_p$; the standard deviation of the sample returns, here denoted by $\hat{\sigma}_p$ (s_p is a familiar notation in other contexts); and the sample mean risk-free rate, $\overline{R}_f$. The *ex post* (or realized or historical) Sharpe ratio can then be determined by using the following:

$$\widehat{\text{SR}} = \frac{\overline{R}_p - \overline{R}_F}{\hat{\sigma}_p}$$

Recalling the CAL from earlier in the reading, one can see that the Sharpe ratio, also called the reward-to-variability ratio, is simply the slope of the capital allocation line. Note, however, that the ratio uses the *total risk* of the portfolio, not its systematic risk. The use of total risk is appropriate if the portfolio is an investor's total portfolio—that is, the investor does not own any other assets. Sharpe ratios of the market and other portfolios can also be calculated in a similar manner. The portfolio with the highest Sharpe ratio has the best risk-adjusted performance, and the one with the lowest Sharpe ratio has the worst risk-adjusted performance, provided that the numerator is positive for all comparison portfolios. If the numerator is negative, the ratio will be less negative for riskier portfolios, resulting in incorrect rankings.

The Sharpe ratio, however, suffers from two limitations. First, it uses total risk as a measure of risk when only systematic risk is priced. Second, the ratio itself (e.g., 0.2 or 0.3) is not informative. To rank portfolios, the Sharpe ratio of one portfolio must be compared with the Sharpe ratio of another portfolio. Nonetheless, the ease of computation makes the Sharpe ratio a popular tool.

11.2 The Treynor Ratio

The **Treynor ratio** is a simple extension of the Sharpe ratio and resolves the Sharpe ratio's first limitation by substituting beta (systematic risk) for total risk. The *ex ante* and *ex post* Treynor ratios are provided below.

$$\text{TR} = \frac{E\left(R_p\right) - R_f}{\beta_p}$$

$$\widehat{\mathrm{TR}} = \frac{\bar{R}_p - \bar{R}_f}{\hat{\beta}_p}$$

Just like the Sharpe ratio, the numerators must be positive for the Treynor ratio to give meaningful results. In addition, the Treynor ratio does not work for negative-beta assets—that is, the denominator must also be positive for obtaining correct estimates and rankings. Although both the Sharpe and Treynor ratios allow for ranking of portfolios, neither ratio gives any information about the economic significance of differences in performance. For example, assume the Sharpe ratio of one portfolio is 0.75 and the Sharpe ratio for another portfolio is 0.80. The second portfolio is superior, but is that difference meaningful? In addition, we do not know whether either of the portfolios is better than the passive market portfolio. The remaining two measures, M^2 and Jensen's alpha, attempt to address that problem by comparing portfolios while also providing information about the extent of the overperformance or underperformance.

11.3 M^2: Risk-Adjusted Performance (RAP)

M^2 provides a measure of portfolio return that is adjusted for the total risk of the portfolio relative to that of some benchmark. In 1997, Nobel Prize winner Franco Modigliani and his granddaughter, Leah Modigliani, developed what they called a risk-adjusted performance measure, or RAP. The RAP measure has since become more commonly known as M^2 reflecting the Modigliani names. It is related to the Sharpe ratio and ranks portfolios identically, but it has the useful advantage of being denominated in familiar terms of percentage return advantage assuming the same level of total risk as the market

M^2 borrows from capital market theory by assuming a portfolio is leveraged or de-leveraged until its volatility (as measured by standard deviation) matches that of the market. This adjustment produces a portfolio-specific leverage ratio that equates the portfolio's risk to that of the market. The portfolio's excess return times the leverage ratio plus the risk-free rate is then compared with the markets actual return to determine whether the portfolio has outperformed or underperformed the market on a risk-adjusted basis.

The equations below provide the *ex ante* and *ex post* formulas for M^2, where σ_m is the standard deviation of the market portfolio and σ_m/σ_p is the portfolio-specific leverage ratio. Because the Sharpe ratio is defined as $\frac{E(R_p) - R_f}{\sigma_p}$, the equation shows that M^2 can be thought of as a rescaling of the Sharpe ratio that allows for easier comparisons among different portfolios. The reason that M^2 and Sharpe ratios rank portfolios identically is because, in a given time period—and for any given comparison of the market portfolio—both the risk-free rate and the market volatility are constant across all comparisons. Only the Sharpe ratio differs, so it determines all rankings.

$$M^2 = \left[E(R_p) - R_f\right]\frac{\sigma_m}{\sigma_p} + R_f = \mathrm{SR} \times \sigma_m + R_f \text{ (ex ante)}$$

$$\widehat{M^2} = \left(\bar{R}_p - \bar{R}_f\right)\frac{\hat{\sigma}_m}{\hat{\sigma}_p} + R_f = \widehat{\mathrm{SR}} \times \hat{\sigma}_m + R_f \text{ (ex post)}$$

For example, assume that $\bar{R}_f = 4.0\%$, $\bar{R}_p = 14.0\%$, $\hat{\sigma}_p = 25.0\%$ and $\hat{\sigma}_m = 20.0\%$. The Sharpe ratio is 0.4, $\widehat{\mathrm{SR}} = \frac{0.14 - 0.04}{0.25} = 0.4$, and $\widehat{M^2}$ is 12.0%, $\widehat{M^2} = 0.4(0.2) + 0.04 = 0.12 = 12.0\%$. If the market return was 10%, then the portfolio outperformed the market on a risk-adjusted basis by 12.0% – 10.0% = 2.0%. This difference between the risk-adjusted performance of the portfolio and the performance of the market is frequently referred to as **M² alpha**.

The Sharpe ratio of the market portfolio is $\widehat{SR} = \dfrac{0.10 - 0.04}{0.20} = 0.3$. Comparing the Sharpe ratio of the portfolio with the Sharpe ratio of the market portfolio shows that the fund outperformed the market. But the 2.0% difference between M^2 and the market's return tells us the risk-adjusted outperformance as a percentage return.

11.4 Jensen's Alpha

Like the Treynor ratio, Jensen's alpha is based on systematic risk. We can measure a portfolio's systematic risk by estimating the market model, which is done by regressing the portfolio's daily return on the market's daily return. The coefficient on the market return is an estimate of the beta risk of the portfolio. We can calculate the risk-adjusted return of the portfolio using the beta of the portfolio and the CAPM. The difference between the actual portfolio return and the calculated risk-adjusted return is a measure of the portfolio's performance relative to the market portfolio and is called Jensen's alpha. By definition, α_m of the market is zero. Jensen's alpha is also the vertical distance from the SML measuring the excess return for the same risk as that of the market and is given by

$$\alpha_p = R_p - \{R_f + \beta_p[E(R_m) - R_f]\}$$

If the period is long, it may contain different risk-free rates, in which case R_f represents the average risk-free rate. Furthermore, the returns in the equation are all realized, actual returns. The sign of α_p indicates whether the portfolio has outperformed the market. If α_p is positive, then the portfolio has outperformed the market; if α_p is negative, the portfolio has underperformed the market. Jensen's alpha is commonly used for evaluating most institutional managers, pension funds, and mutual funds. Values of alpha can be used to rank different managers and the performance of their portfolios, as well as the magnitude of underperformance or overperformance. For example, if a portfolio's alpha is 2 percent and another portfolio's alpha is 5 percent, the second portfolio has outperformed the first portfolio by 3 percentage points and the market by 5 percentage points. Jensen's alpha is the maximum amount that you should be willing to pay the manager to manage your money. As with other performance appraisal measures, Jensen's alpha has *ex ante* and *ex post* forms. The use context usually clarifies which one is being referred to. Where we want to underscore a reference to *ex post* Jensen's alpha based on an estimated beta, $\hat{\beta}_p$, and an average market return, the notation $\hat{\alpha}_p$ is used.

EXAMPLE 10

Portfolio Performance Evaluation

A British pension fund has employed three investment managers, each of whom is responsible for investing in one-third of all asset classes so that the pension fund has a well-diversified portfolio. Information about the managers is given below.

Manager	Average Return	$\hat{\sigma}$	$\hat{\beta}$
X	10%	20%	1.1
Y	11	10	0.7

Manager	Average Return	$\hat{\sigma}$	$\hat{\beta}$
Z	12	25	0.6
Market (M)	9	19	
Risk-free rate (R_f)	3		

Calculate the expected return for each manager, based on using the average market return and the CAPM. Then also calculate for the managers (ex post) Sharpe ratio, Treynor ratio, M^2 alpha, and Jensen's alpha. Analyze your results and plot the returns and betas of these portfolios.

Solution:

In each case, the calculations are shown only for Manager X. All answers are tabulated below. Note that the β of the market is 1 and the σ and β of the risk-free rate are both zero.

$$\text{Expected return: } E(R_X) = R_f + \beta_X\left[E(R_m) - R_f\right] = 0.03 + 1.10 \times (0.09 - 0.03) = 0.096 = 9.6\%$$

$$\widehat{\text{SR}} = \frac{\bar{R}_x - \bar{R}_f}{\hat{\sigma}_x} = \frac{0.10 - 0.03}{0.20} = 0.35$$

$$\widehat{\text{TR}} = \frac{\bar{R}_x - \bar{R}_f}{\hat{\beta}_x} = \frac{0.10 - 0.03}{1.1} = 0.064$$

$$\widehat{M^2} = \left(\bar{R}_x - \bar{R}_f\right)\frac{\hat{\sigma}_m}{\hat{\sigma}_x} + \bar{R}_f = \widehat{SR} \times \hat{\sigma}_m + \bar{R}_f$$
$$= 0.35 \times 0.19 + 0.03 = 0.0965 = 9.65\%$$

Since the market return is 9%, M^2 alpha is 0.65% (9.65% – 9%).

$$\hat{\alpha}_X = R_X - \left[\bar{R}_f + \hat{\beta}_X\left(\bar{R}_m - \bar{R}_f\right)\right] = 0.10 - (0.03 + 1.1 \times 0.06)$$
$$= 0.004 = 0.40\%$$

Exhibit 8 Measures of Portfolio Performance Evaluation

Manager	$\bar{R}_i$	$\hat{\sigma}_i$	$\hat{\beta}_i$	$E(R_i)$	Sharpe Ratio	Treynor Ratio	M^2 alpha	$\hat{\alpha}_i$
X	10.0%	20.0%	1.10	9.6%	0.35	0.064	0.65%	0.40%
Y	11.0	10.0	0.70	7.2	0.80	0.114	9.20	3.80
Z	12.0	25.0	0.60	6.6	0.36	0.150	0.84	5.40
M	9.0	19.0	1.00	9.0	0.32	0.060	0.00	0.00
R_f	3.0	0.0	0.00	3.0	–	–	–	0.00

Let us begin with an analysis of the risk-free asset. Because the risk-free asset has zero risk and a beta of zero, calculating the Sharpe ratio, Treynor ratio, or M^2 is not possible because they all require the portfolio risk in the denominator. The risk-free asset's alpha, however, is zero. Turning to the market portfolio, we see that the absolute measures of performance, the Sharpe ratio and the Treynor ratio, are positive for the market portfolio. These ratios are positive as long as the portfolio earns a return that is in excess of that of the risk-free asset. $\widehat{M}^2$ and $\hat{\alpha}_i$ are performance measures relative to the market, so they are both equal to zero for the market portfolio.

All three managers have Sharpe and Treynor ratios greater than those of the market, and all three managers' M^2 alpha and α_i are positive; therefore, the pension fund should be satisfied with their performance. Among the three managers, Manager X has the worst performance, irrespective of whether total risk or systematic risk is considered for measuring performance. The relative rankings are depicted in Exhibit 9.

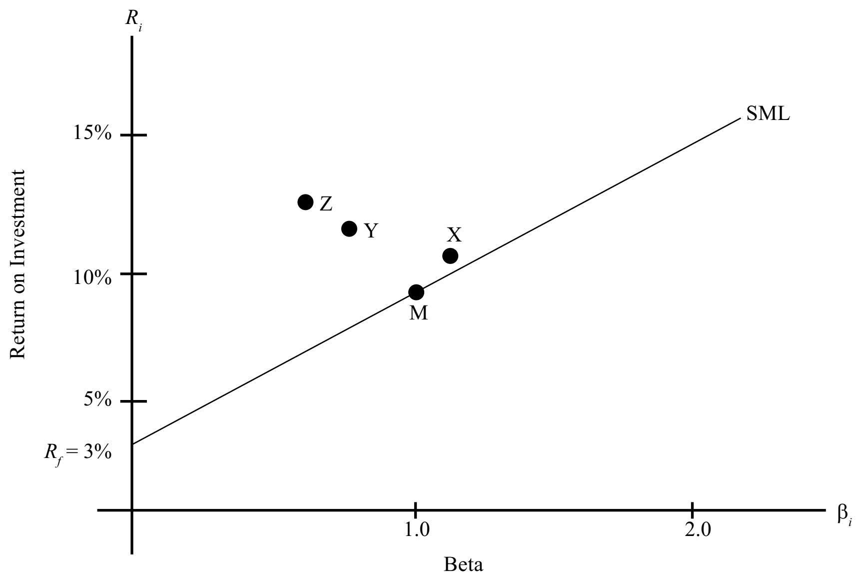
Exhibit 9 Ranking of Portfolios by Performance Measure

Rank	Sharpe Ratio	Treynor Ratio	M^2 alpha	α_i
1	Y	Z	Y	Z
2	Z	Y	Z	Y
3	X	X	X	X
4	M	M	M	M
5	–	–	–	R_f

Comparing Y and Z, we can observe that Y performs much better than Z when total risk is considered. Y has a Sharpe ratio of 0.80, compared with a Sharpe ratio of 0.36 for Z. Similarly, M^2 alpha is higher for Y (9.20 percent) than for Z (0.84 percent). In contrast, when systematic risk is used, Z outperforms Y. The Treynor ratio is higher for Z (0.150) than for Y (0.114), and Jensen's alpha is also higher for Z (5.40 percent) than for Y (3.80 percent), which indicates that Z has done a better job of generating excess return relative to systematic risk than Y.

Exhibit 10 confirms these observations in that all three managers outperform the benchmark because all three points lie above the SML. Among the three portfolios, Z performs the best when we consider risk-adjusted returns because it is the point in Exhibit 10 that is located northwest relative to the portfolios X and Y.

Exhibit 10 Portfolios Along the SML



When do we use total risk performance measures like the Sharpe ratio and M^2, and when do we use beta risk performance measures like the Treynor ratio and Jensen's alpha? Total risk is relevant for an investor when he or she holds a portfolio that is not fully diversified, which is not a desirable portfolio. In such cases, the Sharpe ratio and M^2 are appropriate performance measures. Thus, if the pension fund were to choose only one fund manager to manage all its assets, it should choose Manager Y. Performance measures relative to beta risk—Treynor ratio and Jensen's alpha—are relevant when the investor holds a well-diversified portfolio with negligible diversifiable risk. In other words, if the pension fund is well diversified and only the systematic risk of the portfolio matters, the fund should choose Manager Z.

The measures of performance evaluation assume that the market portfolio is the correct benchmark. As a result, an error in the benchmark may cause the results to be misleading. For example, evaluating a real estate fund against the S&P 500 is incorrect because real estate has different characteristics than equity. In addition to errors in benchmarking, errors could occur in the measurement of risk and return of the market portfolio and the portfolios being evaluated. Finally, many estimates are based on historical data. Any projections based on such estimates assume that this level of performance will continue in the future.

APPLICATIONS OF THE CAPM IN PORTFOLIO CONSTRUCTION 12

g calculate and interpret the expected return of an asset using the CAPM;

h describe and demonstrate applications of the CAPM and the SML;

This section introduces applications of the CAPM in portfolio construction. First, the security characteristic line, which graphically indicates *ex post* Jensen's alpha, is described. If we relax the assumption that investors have the same expectations about risk and return, a positive Jensen's alpha can be interpreted as an indication of superior information or investment ability. The section on security selection covers that possibility. The last section summarizes how the CAPM and related concepts can be applied to portfolio construction.

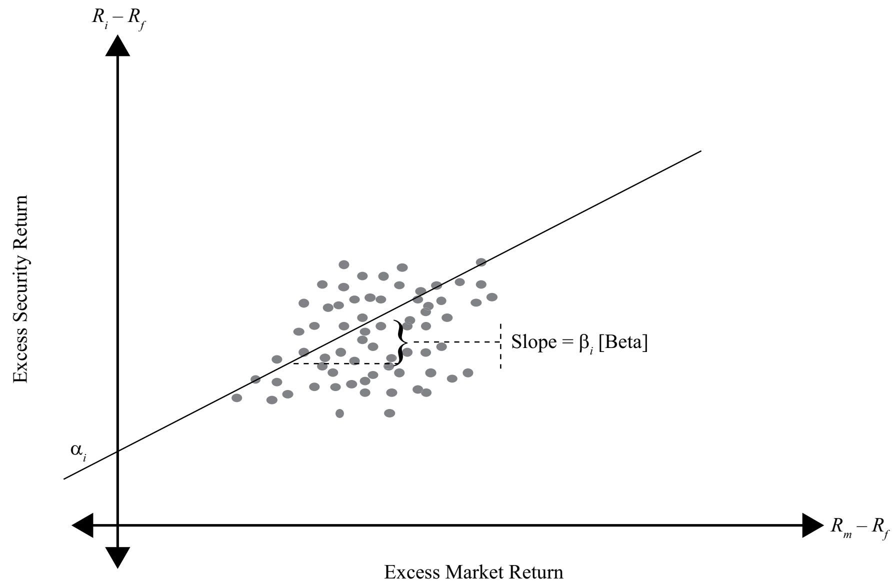
12.1 Security Characteristic Line

Similar to the SML, we can draw a **security characteristic line** (SCL) for a security. The SCL is a plot of the excess return of the security on the excess return of the market. In Exhibit 8, Jensen's alpha is the intercept and the beta is the slope. The equation of the line can be obtained by rearranging the terms in the expression for Jensen's alpha and replacing the subscript *p* with *i*:

$$R_i - R_f = \alpha_i + \beta_i(R_m - R_f)$$

As an example, the SCL is drawn in Exhibit 11 using Manager X's portfolio from Exhibit 8. The security characteristic line can also be estimated by regressing the excess security return, $R_i - R_f$, on the excess market return, $R_m - R_f$.

Exhibit 11 The Security Characteristic Line



12.2 Security Selection

When discussing the CAPM, we assumed that investors have homogeneous expectations and are rational, risk-averse, utility-maximizing investors. With these assumptions, we were able to state that all investors assign the same value to all assets and, therefore, have the same optimal risky portfolio, which is the market portfolio. In other words, we assumed that there is commonality among beliefs about an asset's future cash flows and the required rate of return. Given the required rate of return, we can discount the future cash flows of the asset to arrive at its current value, or price, which is agreed upon by all or most investors.

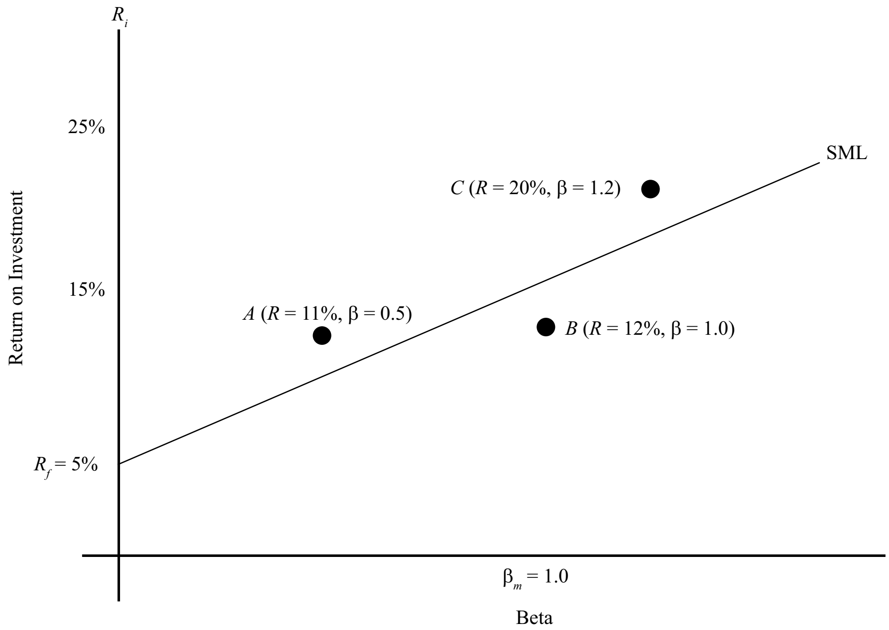
In this section, we introduce heterogeneity in beliefs of investors. Because investors are price takers, it is assumed that such heterogeneity does not significantly affect the market price of an asset. The difference in beliefs can relate to future cash flows, the systematic risk of the asset, or both. Because the current price of an asset is the discounted value of the future cash flows, the difference in beliefs could result in an investor-estimated price that is different from the CAPM-calculated price. The CAPM-calculated price is the current market price because it reflects the beliefs of all other investors in the market. If the investor-estimated current price is higher (lower) than the market price, the asset is considered undervalued (overvalued). Therefore, the CAPM is an effective tool for determining whether an asset is undervalued or overvalued and whether an investor should buy or sell the asset.

Although portfolio performance evaluation is backward looking and security selection is forward looking, we can apply the concepts of portfolio evaluation to security selection. The best measure to apply is Jensen's alpha because it uses systematic risk and is meaningful even on an absolute basis. A positive Jensen's alpha indicates a superior security, whereas a negative Jensen's alpha indicates a security that is likely to underperform the market when adjusted for risk.

Another way of presenting the same information is with the security market line. Potential investors can plot a security's expected return and beta against the SML and use this relationship to decide whether the security is overvalued or undervalued in the market.[6] Exhibit 12 shows a number of securities along with the SML. All securities that reflect the consensus market view are points directly on the SML (i.e., properly valued). If a point representing the estimated return of an asset is above the SML (Points A and C), the asset has a low level of risk relative to the amount of expected return and would be a good choice for investment. In contrast, if the point representing a particular asset is below the SML (Point B), the stock is considered overvalued. Its return does not compensate for the level of risk and should not be considered for investment. Of course, a short position in Asset B can be taken if short selling is permitted.

6 In this reading, we do not consider transaction costs, which are important whenever deviations from a passive portfolio are considered. Thus, the magnitude of undervaluation or overvaluation should be considered in relation to transaction costs prior to making an investment decision.

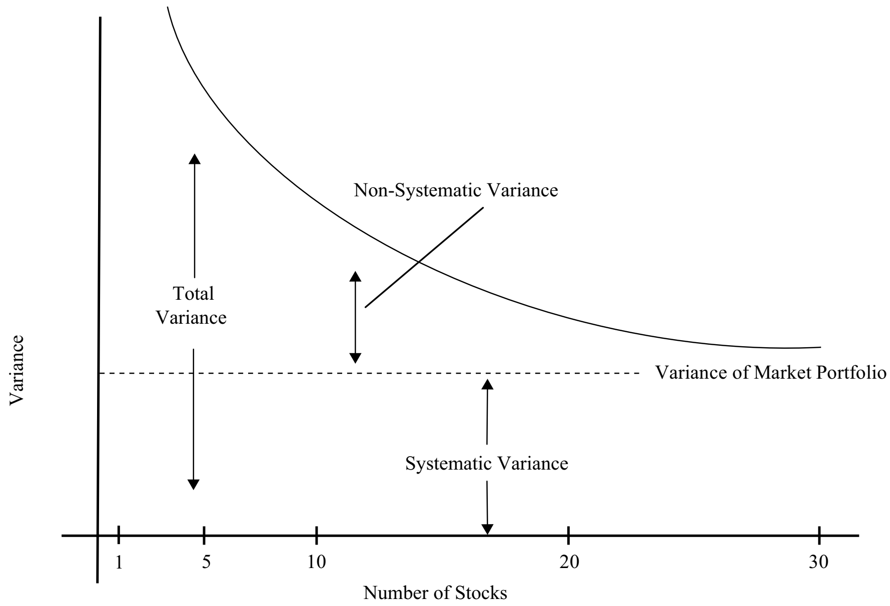
Exhibit 12 Security Selection Using SML



12.3 Implications of the CAPM for Portfolio Construction

Based on the CAPM, investors should hold a combination of the risk-free asset and the market portfolio. The true market portfolio consists of a large number of securities, and an investor would have to own all of them in order to be completely diversified. Because owning all existing securities is not practical, in this section, we will consider an alternate method of constructing a portfolio that may not require a large number of securities and will still be sufficiently diversified. Exhibit 13 shows the reduction in risk as we add more and more securities to a portfolio. As can be seen from the exhibit, much of the nonsystematic risk can be diversified away in as few as 30 securities. These securities, however, should be randomly selected and represent different asset classes for the portfolio to effectively diversify risk. Otherwise, one may be better off using an index (e.g., the S&P 500 for a diversified large-cap equity portfolio and other indexes for other asset classes).

Exhibit 13 Diversification with Number of Stocks



Let's begin constructing the optimal portfolio with a portfolio of securities like the S&P 500. Although the S&P 500 is a portfolio of 500 securities, it is a good starting point because it is readily available as a single security for trading. In contrast, it represents only the large corporations that are traded on the US stock markets and, therefore, does not encompass the global market entirely. Because the S&P 500 is the base portfolio, however, we treat it is as the market for the CAPM.

Any security not included in the S&P 500 can be evaluated to determine whether it should be integrated into the portfolio. That decision is based on the α_i of the security, which is calculated using the CAPM with the S&P 500 as the market portfolio. Note that security i may not necessarily be priced incorrectly for it to have a non-zero α_i; α_i can be positive merely because it is not well correlated with the S&P 500 and its return is sufficient for the amount of systematic risk it contains. For example, assume a new stock market, ABC, opens to foreign investors only and is being considered for inclusion in the portfolio. We estimate ABC's model parameters relative to the S&P 500 and find an α_i of approximately 3 percent, with a β_i of 0.60. Because α_i is positive, ABC should be added to the portfolio. Securities with a significantly negative α_i may be short sold to maximize risk-adjusted return. For convenience, however, we will assume that negative positions are not permitted in the portfolio.

In addition to the securities that are correctly priced but enter the portfolio because of their risk–return superiority, securities already in the portfolio (S&P 500) may be undervalued or overvalued based on investor expectations that are incongruent with the market. Securities in the S&P 500 that are overvalued (negative α_i) should be dropped from the S&P 500 portfolio, if it is possible to exclude individual securities, and positions in securities in the S&P 500 that are undervalued (positive α_i) should be increased.

This brings us to the next question: What should the relative weight of securities in the portfolio be? Because we are concerned with maximizing risk-adjusted return, securities with a higher α_i should have a higher weight, and securities with greater nonsystematic risk should be given less weight in the portfolio. A complete analysis of portfolio optimization is beyond the scope of this reading, but the following principles are helpful. The weight in each nonmarket security should be proportional to $\frac{\alpha_i}{\sigma_{ei}^2}$, where the denominator is the nonsystematic variance of security *i*. The total weight of nonmarket securities in the portfolio is proportional to $\frac{\sum_{i=1}^{N} w_i \alpha_i}{\sum_{i=1}^{N} w_i^2 \sigma_{ei}^2}$. The weight in the market portfolio is a function of $\frac{E(R_m)}{\sigma_m^2}$. The information ratio, $\frac{\alpha_i}{\sigma_{ei}}$ (i.e., alpha divided by nonsystematic risk), measures the abnormal return per unit of risk added by the security to a well-diversified portfolio. The larger the information ratio is, the more valuable the security.

EXAMPLE 11

Optimal Investor Portfolio with Heterogeneous Beliefs

A Japanese investor is holding the Nikkei 225 index, which is her version of the market. She thinks that three stocks, P, Q, and R, which are not in the Nikkei 225, are undervalued and should form a part of her portfolio. She has the following information about the stocks, the Nikkei 225, and the risk-free rate (the information is given as expected return, standard deviation, and beta):

P: 15%, 30%, 1.5

Q: 18%, 25%, 1.2

R: 16%, 23%, 1.1

Nikkei 225: 12%, 18%, 1.0

Risk-free rate: 2%, 0%, 0.0

1 Calculate Jensen's alpha for P, Q, and R.
2 Calculate nonsystematic variance for P, Q, and R.
3 Should any of the three stocks be included in the portfolio? If so, which stock should have the highest weight in the portfolio?

Solution to 1:

Stock P's α: $R_i - [R_f + \beta_i(R_m - R_f)] = 0.15 - (0.02 + 1.5 \times 0.10) = -0.02$

Stock Q's α: $R_i - [R_f + \beta_i(R_m - R_f)] = 0.18 - (0.02 + 1.2 \times 0.10) = 0.04$

Stock R's α: $R_i - [R_f + \beta_i(R_m - R_f)] = 0.16 - (0.02 + 1.1 \times 0.10) = 0.03$

Solution to 2:

Total variance = Systematic variance + Nonsystematic variance. From Section 6, we write the equation as $\sigma_{ei}^2 = \sigma_i^2 - \beta_i^2 \sigma_m^2$.

Stock P's nonsystematic variance = (0.30 × 0.30) – (1.5 × 1.5 × 0.18 × 0.18) = 0.09 – 0.0729 = 0.0171

Stock Q's nonsystematic variance = (0.25 × 0.25) – (1.2 × 1.2 × 0.18 × 0.18) = 0.0625 – 0.0467 = 0.0158

Stock R's nonsystematic variance = (0.23 × 0.23) – (1.1 × 1.1 × 0.18 × 0.18) = 0.0529 – 0.0392 = 0.0137

Solution to 3:

Stock P has a negative α and should not be included in the portfolio, unless a negative position can be assumed through short selling. Stocks Q and R have a positive α; therefore, they should be included in the portfolio with positive weights.

The relative weight of Q is 0.04/0.0158 = 2.53.

The relative weight of R is 0.03/0.0137 = 2.19.

Stock Q will have the largest weight among the nonmarket securities to be added to the portfolio. In relative terms, the weight of Q will be 15.5 percent greater than the weight of R (2.53/2.19 = 1.155). As the number of securities increases, the analysis becomes more complex. However, the contribution of each additional security toward improvement in the risk–return trade-off will decrease and eventually disappear, resulting in a well-diversified portfolio.

SUMMARY

In this reading, we discussed the capital asset pricing model in detail and covered related topics such as the capital market line. The reading began with an interpretation of the CML, uses of the market portfolio as a passive management strategy, and leveraging of the market portfolio to obtain a higher expected return. Next, we discussed systematic and nonsystematic risk and why one should not expect to be compensated for taking on nonsystematic risk. The discussion of systematic and nonsystematic risk was followed by an introduction to beta and return-generating models. This broad topic was then broken down into a discussion of the CAPM and, more specifically, the relationship between beta and expected return. The final section included applications of the CAPM to capital budgeting, portfolio performance evaluation, and security selection. The highlights of the reading are as follows.

- The capital market line is a special case of the capital allocation line, where the efficient portfolio is the market portfolio.
- Obtaining a unique optimal risky portfolio is not possible if investors are permitted to have heterogeneous beliefs because such beliefs will result in heterogeneous asset prices.
- Investors can leverage their portfolios by borrowing money and investing in the market.
- Systematic risk is the risk that affects the entire market or economy and is not diversifiable.
- Nonsystematic risk is local and can be diversified away by combining assets with low correlations.
- Beta risk, or systematic risk, is priced and earns a return, whereas nonsystematic risk is not priced.
- The expected return of an asset depends on its beta risk and can be computed using the CAPM, which is given by $E(R_i) = R_f + \beta_i[E(R_m) - R_f]$.
- The security market line is an implementation of the CAPM and applies to all securities, whether they are efficient or not.
- Expected return from the CAPM can be used for making capital budgeting decisions.

- Portfolios can be evaluated by several CAPM-based measures, such as the Sharpe ratio, the Treynor ratio, M^2, and Jensen's alpha.
- The SML can assist in security selection and optimal portfolio construction.

By successfully understanding the content of this reading, you should feel comfortable decomposing total variance into systematic and nonsystematic variance, analyzing beta risk, using the CAPM, and evaluating portfolios and individual securities.

REFERENCES

Carhart, Mark. 1997. "On Persistence in Mutual Fund Performance." *Journal of Finance*, vol. 52, no. 1:57–82.

Fama, Eugene, and Kenneth French. 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance*, vol. 47, no. 2:427–466.

Lintner, John. 1965a. "Security Prices, Risk, and Maximal Gains from Diversification." *Journal of Finance*, vol. 20, no. 4:587–615.

Lintner, John. 1965b. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics*, vol. 47, no. 1:13–37.

Markowitz, Harry. 1952. "Portfolio Selection." *Journal of Finance*, vol. 7, no. 1:77–91.

Mossin, Jan. 1966. "Equilibrium in a Capital Asset Market." *Econometrica*, vol. 34, no. 4:768–783.

Roll, Richard. 1977. "A Critique of the Asset Pricing Theory's Tests Part I: On Past and Potential Testability of the Theory." *Journal of Financial Economics*, vol. 4, no. 2:129–176.

Ross, Stephen A. 1976. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory*, vol. 13, no. 3:341–360.

Sharpe, William F. 1964. "Capital Asset Prices: A Theory Of Market Equilibrium under Conditions of Risk." *Journal of Finance*, vol. 19, no. 3:425–442.

Treynor, Jack L. 1961. *Market Value, Time, and Risk*. Unpublished manuscript.

Treynor, Jack L. 1962. *Toward a Theory of Market Value of Risky Assets*. Unpublished manuscript.

PRACTICE PROBLEMS

1 The line depicting the total risk and expected return of portfolio combinations of a risk-free asset and any risky asset is the:
   - A security market line.
   - B capital allocation line.
   - C security characteristic line.

2 The portfolio of a risk-free asset and a risky asset has a better risk-return tradeoff than investing in only one asset type because the correlation between the risk-free asset and the risky asset is equal to:
   - A −1.0.
   - B 0.0.
   - C 1.0.

3 With respect to capital market theory, an investor's optimal portfolio is the combination of a risk-free asset and a risky asset with the highest:
   - A expected return.
   - B indifference curve.
   - C capital allocation line slope.

4 Highly risk-averse investors will *most likely* invest the majority of their wealth in:
   - A risky assets.
   - B risk-free assets.
   - C the optimal risky portfolio.

5 The capital market line (CML) is the graph of the risk and return of portfolio combinations consisting of the risk-free asset and:
   - A any risky portfolio.
   - B the market portfolio.
   - C the leveraged portfolio.

6 Which of the following statements *most accurately* defines the market portfolio in capital market theory? The market portfolio consists of all:
   - A risky assets.
   - B tradable assets.
   - C investable assets.

7 With respect to capital market theory, the optimal risky portfolio:
   - A is the market portfolio.
   - B has the highest expected return.
   - C has the lowest expected variance.

8 Relative to portfolios on the CML, any portfolio that plots above the CML is considered:
   - A inferior.
   - B inefficient.
   - C unachievable.

9 A portfolio on the capital market line with returns greater than the returns on the market portfolio represents a(n):

A lending portfolio.

B borrowing portfolio.

C unachievable portfolio.

10 With respect to the capital market line, a portfolio on the CML with returns less than the returns on the market portfolio represents a(n):

A lending portfolio.

B borrowing portfolio.

C unachievable portfolio.

11 Which of the following types of risk is *most likely* avoided by forming a diversified portfolio?

A Total risk.

B Systematic risk.

C Nonsystematic risk.

12 Which of the following events is *most likely* an example of nonsystematic risk?

A A decline in interest rates.

B The resignation of chief executive officer.

C An increase in the value of the US dollar.

13 With respect to the pricing of risk in capital market theory, which of the following statements is *most accurate*?

A All risk is priced.

B Systematic risk is priced.

C Nonsystematic risk is priced.

14 The sum of an asset's systematic variance and its nonsystematic variance of returns is equal to the asset's:

A beta.

B total risk.

C total variance.

15 With respect to return-generating models, the intercept term of the market model is the asset's estimated:

A beta.

B alpha.

C variance.

16 With respect to return-generating models, the slope term of the market model is an estimate of the asset's:

A total risk.

B systematic risk.

C nonsystematic risk.

17 With respect to return-generating models, which of the following statements is *most accurate*? Return-generating models are used to directly estimate the:

A expected return of a security.

B weights of securities in a portfolio.

C parameters of the capital market line.

The following information relates to Questions 18–20

An analyst gathers the following information:

Security	Expected Annual Return (%)	Expected Standard Deviation (%)	Correlation between Security and the Market
Security 1	11	25	0.6
Security 2	11	20	0.7
Security 3	14	20	0.8
Market	10	15	1.0

18 Which security has the *highest* total risk?

- **A** Security 1.
- **B** Security 2.
- **C** Security 3.

19 Which security has the *highest* beta measure?

- **A** Security 1.
- **B** Security 2.
- **C** Security 3.

20 Which security has the *least* amount of market risk?

- **A** Security 1.
- **B** Security 2.
- **C** Security 3.

---

21 With respect to capital market theory, the average beta of all assets in the market is:

- **A** less than 1.0.
- **B** equal to 1.0.
- **C** greater than 1.0.

22 The slope of the security characteristic line is an asset's:

- **A** beta.
- **B** excess return.
- **C** risk premium.

23 The graph of the capital asset pricing model is the:

- **A** capital market line.
- **B** security market line.
- **C** security characteristic line.

24 With respect to capital market theory, correctly priced individual assets can be plotted on the:

- **A** capital market line.
- **B** security market line.
- **C** capital allocation line.

25 With respect to the capital asset pricing model, the primary determinant of expected return of an individual asset is the:

A asset's beta.

B market risk premium.

C asset's standard deviation.

26 With respect to the capital asset pricing model, which of the following values of beta for an asset is *most likely* to have an expected return for the asset that is less than the risk-free rate?

A −0.5

B 0.0

C 0.5

27 With respect to the capital asset pricing model, the market risk premium is:

A less than the excess market return.

B equal to the excess market return.

C greater than the excess market return.

The following information relates to Questions 28–31

An analyst gathers the following information:

Security	Expected Standard Deviation (%)	Beta
Security 1	25	1.50
Security 2	15	1.40
Security 3	20	1.60

28 With respect to the capital asset pricing model, if the expected market risk premium is 6% and the risk-free rate is 3%, the expected return for Security 1 is *closest* to:

A 9.0%.

B 12.0%.

C 13.5%.

29 With respect to the capital asset pricing model, if expected return for Security 2 is equal to 11.4% and the risk-free rate is 3%, the expected return for the market is *closest* to:

A 8.4%.

B 9.0%.

C 10.3%.

30 With respect to the capital asset pricing model, if the expected market risk premium is 6% the security with the *highest* expected return is:

A Security 1.

B Security 2.

C Security 3.

31 With respect to the capital asset pricing model, a decline in the expected market return will have the *greatest* impact on the expected return of:

A Security 1.
B Security 2.
C Security 3.

---

32 Three equity fund managers have performance records summarized in the following table:

	Mean Annual Return (%)	Standard Deviation of Return (%)
Manager 1	14.38	10.53
Manager 2	9.25	6.35
Manager 3	13.10	8.23

Given a risk-free rate of return of 2.60%, which manager performed best based on the Sharpe ratio?

A Manager 1
B Manager 2
C Manager 3

33 Which of the following performance measures is consistent with the CAPM?

A *M*-squared.
B Sharpe ratio.
C Jensen's alpha.

34 Which of the following performance measures does *not* require the measure to be compared to another value?

A Sharpe ratio.
B Treynor ratio.
C Jensen's alpha.

35 Which of the following performance measures is *most* appropriate for an investor who is *not* fully diversified?

A *M*-squared.
B Treynor ratio.
C Jensen's alpha.

36 Analysts who have estimated returns of an asset to be greater than the expected returns generated by the capital asset pricing model should consider the asset to be:

A overvalued.
B undervalued.
C properly valued.

37 With respect to capital market theory, which of the following statements *best* describes the effect of the homogeneity assumption? Because all investors have the same economic expectations of future cash flows for all assets, investors will invest in:

A the same optimal risky portfolio.
B the Standard and Poor's 500 Index.
C assets with the same amount of risk.

38 With respect to capital market theory, which of the following assumptions allows for the existence of the market portfolio? All investors:

A are price takers.

B have homogeneous expectations.

C plan for the same, single holding period.

39 The intercept of the best fit line formed by plotting the excess returns of a manager's portfolio on the excess returns of the market is *best* described as Jensen's:

A beta.

B ratio.

C alpha.

40 Portfolio managers who are maximizing risk-adjusted returns will seek to invest *more* in securities with:

A lower values of Jensen's alpha.

B values of Jensen's alpha equal to 0.

C higher values of Jensen's alpha.

41 Portfolio managers, who are maximizing risk-adjusted returns, will seek to invest *less* in securities with:

A lower values for nonsystematic variance.

B values of nonsystematic variance equal to 0.

C higher values for nonsystematic variance.

SOLUTIONS

1 B is correct. A capital allocation line (CAL) plots the expected return and total risk of combinations of the risk-free asset and a risky asset (or a portfolio of risky assets).

2 B is correct. A portfolio of the risk-free asset and a risky asset or a portfolio of risky assets can result in a better risk-return tradeoff than an investment in only one type of an asset, because the risk-free asset has zero correlation with the risky asset.

3 B is correct. Investors will have different optimal portfolios depending on their indifference curves. The optimal portfolio for each investor is the one with highest utility; that is, where the CAL is tangent to the individual investor's highest possible indifference curve.

4 B is correct. Although the optimal risky portfolio is the market portfolio, highly risk-averse investors choose to invest most of their wealth in the risk-free asset.

5 B is correct. Although the capital allocation line includes all possible combinations of the risk-free asset and any risky portfolio, the capital market line is a special case of the capital allocation line, which uses the market portfolio as the optimal risky portfolio.

6 A is correct. The market includes all risky assets, or anything that has value; however, not all assets are tradable, and not all tradable assets are investable.

7 A is correct. The optimal risky portfolio is the market portfolio. Capital market theory assumes that investors have homogeneous expectations, which means that all investors analyze securities in the same way and are rational. That is, investors use the same probability distributions, use the same inputs for future cash flows, and arrive at the same valuations. Because their valuations of all assets are identical, all investors will invest in the same optimal risky portfolio (i.e., the market portfolio).

8 C is correct. Theoretically, any point above the CML is not achievable and any point below the CML is dominated by and inferior to any point on the CML.

9 B is correct. As one moves further to the right of point M on the capital market line, an increasing amount of borrowed money is being invested in the market portfolio. This means that there is negative investment in the risk-free asset, which is referred to as a leveraged position in the risky portfolio.

10 A is correct. The combinations of the risk-free asset and the market portfolio on the CML where returns are less than the returns on the market portfolio are termed 'lending' portfolios.

11 C is correct. Investors are capable of avoiding nonsystematic risk by forming a portfolio of assets that are not highly correlated with one another, thereby reducing total risk and being exposed only to systematic risk.

12 B is correct. Nonsystematic risk is specific to a firm, whereas systematic risk affects the entire economy.

13 B is correct. Only systematic risk is priced. Investors do not receive any return for accepting nonsystematic or diversifiable risk.

14 C is correct. The sum of systematic variance and nonsystematic variance equals the total variance of the asset. References to total risk as the sum of systematic risk and nonsystematic risk refer to variance, not to risk.

15 B is correct. In the market model, $R_i = \alpha_i + \beta_i R_m + e_i$, the intercept, α_i, and slope coefficient, β_i, are estimated using historical security and market returns.

16 B is correct. In the market model, $R_i = \alpha_i + \beta_i R_m + e_i$, the slope coefficient, β_i, is an estimate of the asset's systematic or market risk.

17 A is correct. In the market model, $R_i = \alpha_i + \beta_i R_m + e_i$, the intercept, α_i, and slope coefficient, β_i, are estimated using historical security and market returns. These parameter estimates then are used to predict firm-specific returns that a security may earn in a future period.

18 A is correct. Security 1 has the highest total risk = 0.25 compared to Security 2 and Security 3 with a total risk of 0.20.

19 C is correct. Security 3 has the highest beta value; $1.07 = \dfrac{\rho_{3,m}\sigma_3}{\sigma_m} = \dfrac{(0.80)(20\%)}{15\%}$ compared to Security 1 and Security 2 with beta values of 1.00 and 0.93, respectively.

20 B is correct. Security 2 has the lowest beta value; $0.93 = \dfrac{\rho_{2,m}\sigma_2}{\sigma_m} = \dfrac{(0.70)(20\%)}{15\%}$ compared to Security 1 and 3 with beta values of 1.00 and 1.07, respectively.

21 B is correct. The average beta of all assets in the market, by definition, is equal to 1.0.

22 A is correct. The security characteristic line is a plot of the excess return of the security on the excess return of the market. In such a graph, Jensen's alpha is the intercept and the beta is the slope.

23 B is correct. The security market line (SML) is a graphical representation of the capital asset pricing model, with beta risk on the x-axis and expected return on the y-axis.

24 B is correct. The security market line applies to any security, efficient or not. The CAL and the CML use the total risk of the asset (or portfolio of assets) rather than its systematic risk, which is the only risk that is priced.

25 A is correct. The CAPM shows that the primary determinant of expected return for an individual asset is its beta, or how well the asset correlates with the market.

26 A is correct. If an asset's beta is negative, the required return will be less than the risk-free rate in the CAPM. When combined with a positive market return, the asset reduces the risk of the overall portfolio, which makes the asset very valuable. Insurance is an example of a negative beta asset.

27 B is correct. In the CAPM, the market risk premium is the difference between the return on the market and the risk-free rate, which is the same as the return in excess of the market return.

28 B is correct. The expected return of Security 1, using the CAPM, is 12.0% = 3% + 1.5(6%); $E(R_i) = R_f + \beta_i[E(R_m) - R_f]$.

29 B is correct. The expected risk premium for Security 2 is 8.4%, (11.4% – 3%), indicates that the expected market risk premium is 6%; therefore, since the risk-free rate is 3% the expected rate of return for the market is 9%. That is, using the CAPM, $E(R_i) = R_f + \beta_i[E(R_m) - R_f]$, 11.4% = 3% + 1.4(X%), where X% = (11.4% – 3%)/1.4 = 6.0% = market risk premium.

30 C is correct. Security 3 has the highest beta; thus, regardless of the value for the risk-free rate, Security 3 will have the highest expected return:

$$E(R_i) = R_f + \beta_i[E(R_m) - R_f]$$

31 C is correct. Security 3 has the highest beta; thus, regardless of the risk-free rate the expected return of Security 3 will be most sensitive to a change in the expected market return.

32 C is correct. The Sharpe ratio $\left(\widehat{\text{SR}}\right)$ is the mean excess portfolio return per unit of risk, where a higher Sharpe ratio indicates better performance:

$$\widehat{SR}_1 = \frac{\overline{R}_p - \overline{R}_f}{\hat{\sigma}_p} = \frac{14.38 - 2.60}{10.53} = 1.12$$

$$\widehat{SR}_2 = \frac{\overline{R}_p - \overline{R}_f}{\hat{\sigma}_p} = \frac{9.25 - 2.60}{6.35} = 1.05$$

$$\widehat{SR}_3 = \frac{\overline{R}_p - \overline{R}_f}{\hat{\sigma}_p} = \frac{13.10 - 2.60}{8.23} = 1.28$$

33 C is correct. Jensen's alpha adjusts for systematic risk, and *M*-squared and the Sharpe Ratio adjust for total risk.

34 C is correct. The sign of Jensen's alpha indicates whether or not the portfolio has outperformed the market. If alpha is positive, the portfolio has outperformed the market; if alpha is negative, the portfolio has underperformed the market.

35 A is the correct. *M*-squared adjusts for risk using standard deviation (i.e., total risk).

36 B is correct. If the estimated return of an asset is above the SML (the expected return), the asset has a lower level of risk relative to the amount of expected return and would be a good choice for investment (i.e., undervalued).

37 A is correct. The homogeneity assumption refers to all investors having the same economic expectation of future cash flows. If all investors have the same expectations, then all investors should invest in the same optimal risky portfolio, therefore implying the existence of only one optimal portfolio (i.e., the market portfolio).

38 B is correct. The homogeneous expectations assumption means that all investors analyze securities in the same way and are rational. That is, they use the same probability distributions, use the same inputs for future cash flows, and arrive at the same valuations. Because their valuation of all assets is identical, they will generate the same optimal risky portfolio, which is the market portfolio.

39 C is correct. This is because of the plot of the excess return of the security on the excess return of the market. In such a graph, Jensen's alpha is the intercept and the beta is the slope.

40 C is correct. Since managers are concerned with maximizing risk-adjusted returns, securities with a higher value of Jensen's alpha, α_i, should have a higher weight.

41 C is correct. Since managers are concerned with maximizing risk-adjusted returns, securities with greater nonsystematic risk should be given less weight in the portfolio.