2025 CFA®
Exam Prep

SchweserNotes™
Quantitative Methods, Economics, and Corporate Issuers

Level I Book 1

KAPLAN SCHWESER

Kaplan Schweser's Path to Success

Level I CFA® Exam

CFA®

Welcome

As the head of Advanced Designations at Kaplan Schweser, I am pleased to have the opportunity to help you prepare for the CFA® exam. Kaplan Schweser has decades of experience in delivering the most effective CFA exam prep products in the market and I know you will find them to be invaluable in your studies.

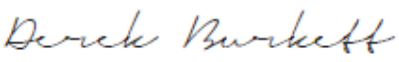
Our products are designed to be an integrated study solution across print and digital media to provide you the best learning experience, whether you are studying with a physical book, online, or on your mobile device.

Our core product, the SchweserNotes™, addresses all of the Topics, Learning Modules, and LOS in the CFA curriculum. Each reading in the SchweserNotes has been broken into smaller, bite-sized modules with Module Quizzes interspersed throughout to help you continually assess your comprehension. Online Topic Quizzes appear at the end of each Topic to help you assess your knowledge of the material before you move on to the next section.

All purchasers of the SchweserNotes receive online access to the Kaplan Schweser online platform (our learning management system or LMS) at www.Schweser.com. In the LMS, you will see a dashboard that tracks your overall progress and performance and also includes an Activity Feed, which provides structure and organization to the tasks required to prepare for the CFA exam. You also have access to the SchweserNotes, Module Quizzes, and Topic Quizzes content as well as the Module Videos (if purchased), which contain a short video that complements each module in the SchweserNotes. Look for the icons indicating where video content, Module Quizzes, and Topic Quizzes are available online. I strongly encourage you to use the dashboard to track your progress and stay motivated.

Again, thank you for trusting Kaplan Schweser with your CFA exam preparation. We're here to help you throughout your journey to become a CFA charterholder.

Regards,



Derek Burkett, CFA, FRM, CAIA
Vice President (Advanced Designations)

Contact us for questions about your study package, upgrading your package, purchasing additional study materials, or for additional information:

888.325.5072 (U.S.) | +1 608.779.8327 (Int'l.)

staff@schweser.com | www.schweser.com/cfa

Book 1: Quantitative Methods, Economics, and Corporate Issuers

SchweserNotes™ 2025

Level I CFA®



SCHWESERNOTES™ 2025 LEVEL I CFA® BOOK 1: QUANTITATIVE METHODS, ECONOMICS, AND CORPORATE ISSUERS

©2024 Kaplan, Inc. All rights reserved.
Published in 2024 by Kaplan, Inc.

ISBN: 978-1-0788-4560-1

These materials may not be copied without written permission from the author. The unauthorized duplication of these notes is a violation of global copyright laws and the CFA Institute Code of Ethics. Your assistance in pursuing potential violators of this law is greatly appreciated.

Required CFA Institute Disclaimer: "Kaplan Schweser is a CFA Institute Prep Provider. Only CFA Institute Prep Providers are permitted to make use of CFA Institute copyrighted materials which are the building blocks of the exam. We are also required to create/use updated materials every year and this is validated by CFA Institute. Our products and services substantially cover the relevant curriculum and exam and this is validated by CFA Institute. In our advertising, any statement about the numbers of questions in our products and services relates to unique, original, proprietary questions. CFA Institute Prep Providers are forbidden from including CFA Institute official mock exam questions or any questions other than the end of reading questions within their products and services.

CFA Institute does not endorse, promote, review or warrant the accuracy or quality of the product and services offered by Kaplan Schweser. CFA Institute®, CFA® and "Chartered Financial Analyst®" are trademarks owned by CFA Institute."

Certain materials contained within this text are the copyrighted property of CFA Institute. The following is the copyright disclosure for these materials: "© Copyright CFA Institute".

Disclaimer: The Schweser study tools should be used in conjunction with the original readings as set forth by CFA Institute. The information contained in these study tools covers topics contained in the readings referenced by CFA Institute and is believed to be accurate. However, their accuracy cannot be guaranteed nor is any warranty conveyed as to your ultimate exam success. The authors of the referenced readings have not endorsed or sponsored these study tools.

CONTENTS

Learning Outcome Statements (LOS)

QUANTITATIVE METHODS

READING 1
Rates and Returns

Module 1.1: Interest Rates and Return Measurement
Module 1.2: Time-Weighted and Money-Weighted Returns
Module 1.3: Common Measures of Return
Key Concepts
Answer Key for Module Quizzes

READING 2
The Time Value of Money in Finance

Module 2.1: Discounted Cash Flow Valuation
Module 2.2: Implied Returns and Cash Flow Additivity
Key Concepts
Answer Key for Module Quizzes

READING 3
Statistical Measures of Asset Returns

Module 3.1: Central Tendency and Dispersion
Module 3.2: Skewness, Kurtosis, and Correlation
Key Concepts
Answer Key for Module Quizzes

READING 4
Probability Trees and Conditional Expectations

Module 4.1: Probability Models, Expected Values, and Bayes' Formula
Key Concepts
Answer Key for Module Quizzes

READING 5
Portfolio Mathematics

Module 5.1: Probability Models for Portfolio Return and Risk
Key Concepts
Answer Key for Module Quizzes

READING 6
Simulation Methods

Module 6.1: Lognormal Distributions and Simulation Techniques
Key Concepts
Answer Key for Module Quizzes

READING 7
Estimation and Inference

Module 7.1: Sampling Techniques and the Central Limit Theorem
Key Concepts
Answer Key for Module Quizzes

READING 8
Hypothesis Testing

Module 8.1: Hypothesis Testing Basics
Module 8.2: Types of Hypothesis Tests
Key Concepts
Answer Key for Module Quizzes

READING 9
Parametric and Non-Parametric Tests of Independence

Module 9.1: Tests for Independence
Key Concepts
Answer Key for Module Quizzes

READING 10
Simple Linear Regression

Module 10.1: Linear Regression Basics
Module 10.2: Analysis of Variance (ANOVA) and Goodness of Fit
Module 10.3: Predicted Values and Functional Forms of Regression
Key Concepts
Answer Key for Module Quizzes

READING 11
Introduction to Big Data Techniques

Module 11.1: Introduction to Fintech
Key Concepts
Answer Key for Module Quizzes

Topic Quiz: Quantitative Methods

ECONOMICS

READING 12
Firms and Market Structures

Module 12.1: Breakeven, Shutdown, and Scale
Module 12.2: Characteristics of Market Structures

Module 12.3: Identifying Market Structures
Key Concepts
Answer Key for Module Quizzes

READING 13
Understanding Business Cycles

Module 13.1: Business Cycles
Key Concepts
Answer Key for Module Quizzes

READING 14
Fiscal Policy

Module 14.1: Fiscal Policy Objectives
Module 14.2: Fiscal Policy Tools and Implementation
Key Concepts
Answer Key for Module Quizzes

READING 15
Monetary Policy

Module 15.1: Central Bank Objectives and Tools
Module 15.2: Monetary Policy Effects and Limitations
Key Concepts
Answer Key for Module Quizzes

READING 16
Introduction to Geopolitics

Module 16.1: Geopolitics
Key Concepts
Answer Key for Module Quizzes

READING 17
International Trade

Module 17.1: International Trade
Key Concepts
Answer Key for Module Quizzes

READING 18
Capital Flows and the FX Market

Module 18.1: The Foreign Exchange Market
Module 18.2: Managing Exchange Rates
Key Concepts
Answer Key for Module Quizzes

READING 19
Exchange Rate Calculations

Module 19.1: Foreign Exchange Rates
Key Concepts
Answer Key for Module Quizzes

Topic Quiz: Economics

CORPORATE ISSUERS

READING 20
Organizational Forms, Corporate Issuer Features, and Ownership

Module 20.1: Features of Corporate Issuers
Key Concepts
Answer Key for Module Quizzes

READING 21
Investors and Other Stakeholders

Module 21.1: Stakeholders and ESG Factors
Key Concepts
Answer Key for Module Quizzes

READING 22
Corporate Governance: Conflicts, Mechanisms, Risks, and Benefits

Module 22.1: Corporate Governance
Key Concepts
Answer Key for Module Quizzes

READING 23
Working Capital and Liquidity

Module 23.1: Liquidity Measures and Management
Key Concepts
Answer Key for Module Quizzes

READING 24
Capital Investments and Capital Allocation

Module 24.1: Capital Investments and Project Measures
Module 24.2: Capital Allocation Principles and Real Options
Key Concepts
Answer Key for Module Quizzes

READING 25
Capital Structure

Module 25.1: Weighted-Average Cost of Capital
Module 25.2: Capital Structure Theories
Key Concepts
Answer Key for Module Quizzes

READING 26

Business Models

Module 26.1: Business Model Features and Types
Key Concepts
Answer Key for Module Quizzes

Topic Quiz: Corporate Issuers

Formulas

Appendix A: Areas Under the Normal Curve
Cumulative *Z*-Table
Appendix B: Student's *t*-Distribution
Appendix C: *F*-Table at 5% (Upper Tail)
Appendix D: *F*-Table at 2.5% (Upper Tail)
Appendix E: Chi-Squared table

Index

Learning Outcome Statements (LOS)

1. Rates and Returns

The candidate should be able to:

a. interpret interest rates as required rates of return, discount rates, or opportunity costs and explain an interest rate as the sum of a real risk-free rate and premiums that compensate investors for bearing distinct types of risk.
b. calculate and interpret different approaches to return measurement over time and describe their appropriate uses.
c. compare the money-weighted and time-weighted rates of return and evaluate the performance of portfolios based on these measures.
d. calculate and interpret annualized return measures and continuously compounded returns, and describe their appropriate uses.
e. calculate and interpret major return measures and describe their appropriate uses.

2. The Time Value of Money in Finance

The candidate should be able to:

a. calculate and interpret the present value (PV) of fixed-income and equity instruments based on expected future cash flows.
b. calculate and interpret the implied return of fixed-income instruments and required return and implied growth of equity instruments given the present value (PV) and cash flows.
c. explain the cash flow additivity principle, its importance for the no-arbitrage condition, and its use in calculating implied forward interest rates, forward exchange rates, and option values.

3. Statistical Measures of Asset Returns

The candidate should be able to:

a. calculate, interpret, and evaluate measures of central tendency and location to address an investment problem.
b. calculate, interpret, and evaluate measures of dispersion to address an investment problem.
c. interpret and evaluate measures of skewness and kurtosis to address an investment problem.
d. interpret correlation between two variables to address an investment problem.

4. Probability Trees and Conditional Expectations

The candidate should be able to:

a. calculate expected values, variances, and standard deviations and demonstrate their application to investment problems.
b. formulate an investment problem as a probability tree and explain the use of conditional expectations in investment application.
c. calculate and interpret an updated probability in an investment setting using Bayes' formula.

5. Portfolio Mathematics

The candidate should be able to:

a. calculate and interpret the expected value, variance, standard deviation, covariances, and correlations of portfolio returns.
b. calculate and interpret the covariance and correlation of portfolio returns using a joint probability function for returns.
c. define shortfall risk, calculate the safety-first ratio, and identify an optimal portfolio using Roy's safety-first criterion.

6. Simulation Methods

The candidate should be able to:

a. explain the relationship between normal and lognormal distributions and why the lognormal distribution is used to model asset prices when using continuously compounded asset returns.
b. describe Monte Carlo simulation and explain how it can be used in investment applications.

c. describe the use of bootstrap resampling in conducting a simulation based on observed data in investment applications.

7. Estimation and Inference

The candidate should be able to:

a. compare and contrast simple random, stratified random, cluster, convenience, and judgmental sampling and their implications for sampling error in an investment problem.
b. explain the central limit theorem and its importance for the distribution and standard error of the sample mean.
c. describe the use of resampling (bootstrap, jackknife) to estimate the sampling distribution of a statistic.

8. Hypothesis Testing

The candidate should be able to:

a. explain hypothesis testing and its components, including statistical significance, Type I and Type II errors, and the power of a test.
b. construct hypothesis tests and determine their statistical significance, the associated Type I and Type II errors, and power of the test given a significance level.
c. compare and contrast parametric and nonparametric tests, and describe situations where each is the more appropriate type of test.

9. Parametric and Non-Parametric Tests of Independence

The candidate should be able to:

a. explain parametric and nonparametric tests of the hypothesis that the population correlation coefficient equals zero, and determine whether the hypothesis is rejected at a given level of significance.
b. explain tests of independence based on contingency table data.

10. Simple Linear Regression

The candidate should be able to:

a. describe a simple linear regression model, how the least squares criterion is used to estimate regression coefficients, and the interpretation of these coefficients.
b. explain the assumptions underlying the simple linear regression model, and describe how residuals and residual plots indicate if these assumptions may have been violated.
c. calculate and interpret measures of fit and formulate and evaluate tests of fit and of regression coefficients in a simple linear regression.
d. describe the use of analysis of variance (ANOVA) in regression analysis, interpret ANOVA results, and calculate and interpret the standard error of estimate in a simple linear regression.
e. calculate and interpret the predicted value for the dependent variable, and a prediction interval for it, given an estimated linear regression model and a value for the independent variable.
f. describe different functional forms of simple linear regressions.

11. Introduction to Big Data Techniques

The candidate should be able to:

a. describe aspects of "fintech" that are directly relevant for the gathering and analyzing of financial data.
b. describe Big Data, artificial intelligence, and machine learning.
c. describe applications of Big Data and Data Science to investment management.

12. Firms and Market Structures

The candidate should be able to:

a. determine and interpret breakeven and shutdown points of production, as well as how economies and diseconomies of scale affect costs under perfect and imperfect competition.
b. describe characteristics of perfect competition, monopolistic competition, oligopoly, and pure monopoly.
c. explain supply and demand relationships under monopolistic competition, including the optimal price and output for firms as well as pricing strategy.
d. explain supply and demand relationships under oligopoly, including the optimal price and output for firms as well as pricing strategy.

e. identify the type of market structure within which a firm operates and describe the use and limitations of concentration measures.

13. Understanding Business Cycles

The candidate should be able to:

a. describe the business cycle and its phases.
b. describe credit cycles.
c. describe how resource use, consumer and business activity, housing sector activity, and external trade sector activity vary over the business cycle and describe their measurement using economic indicators.

14. Fiscal Policy

The candidate should be able to:

a. compare monetary and fiscal policy.
b. describe roles and objectives of fiscal policy as well as arguments as to whether the size of a national debt relative to GDP matters.
c. describe tools of fiscal policy, including their advantages and disadvantages.
d. explain the implementation of fiscal policy and difficulties of implementation as well as whether a fiscal policy is expansionary or contractionary.

15. Monetary Policy

The candidate should be able to:

a. describe the roles and objectives of central banks.
b. describe tools used to implement monetary policy tools and the monetary transmission mechanism, and explain the relationships between monetary policy and economic growth, inflation, interest, and exchange rates.
c. describe qualities of effective central banks; contrast their use of inflation, interest rate, and exchange rate targeting in expansionary or contractionary monetary policy; and describe the limitations of monetary policy.
d. explain the interaction of monetary and fiscal policy.

16. Introduction to Geopolitics

The candidate should be able to:

a. describe geopolitics from a cooperation versus competition perspective.
b. describe geopolitics and its relationship with globalization.
c. describe functions and objectives of the international organizations that facilitate trade, including the World Bank, the International Monetary Fund, and the World Trade Organization.
d. describe geopolitical risk.
e. describe tools of geopolitics and their impact on regions and economies.
f. describe the impact of geopolitical risk on investments.

17. International Trade

The candidate should be able to:

a. describe the benefits and costs of international trade.
b. compare types of trade restrictions, such as tariffs, quotas, and export subsidies, and their economic implications.
c. explain motivations for and advantages of trading blocs, common markets, and economic unions.

18. Capital Flows and the FX Market

The candidate should be able to:

a. describe the foreign exchange market, including its functions and participants, distinguish between nominal and real exchange rates, and calculate and interpret the percentage change in a currency relative to another currency.
b. describe exchange rate regimes and explain the effects of exchange rates on countries' international trade and capital flows.
c. describe common objectives of capital restrictions imposed by governments.

19. Exchange Rate Calculations

The candidate should be able to:

a. calculate and interpret currency cross-rates.
b. explain the arbitrage relationship between spot and forward exchange rates and interest rates, calculate a forward rate using points or in percentage terms, and interpret a forward discount or premium.

20. Organizational Forms, Corporate Issuer Features, and Ownership

The candidate should be able to:

a. compare the organizational forms of businesses.
b. describe key features of corporate issuers.
c. compare publicly and privately owned corporate issuers.

21. Investors and Other Stakeholders

The candidate should be able to:

a. compare the financial claims and motivations of lenders and shareholders.
b. describe a company's stakeholder groups and compare their interests.
c. describe environmental, social, and governance factors of corporate issuers considered by investors.

22. Corporate Governance: Conflicts, Mechanisms, Risks, and Benefits

The candidate should be able to:

a. describe the principal-agent relationship and conflicts that may arise between stakeholder groups.
b. describe corporate governance and mechanisms to manage stakeholder relationships and mitigate associated risks.
c. describe potential risks of poor corporate governance and stakeholder management and benefits of effective corporate governance and stakeholder management.

23. Working Capital and Liquidity

The candidate should be able to:

a. explain the cash conversion cycle and compare issuers' cash conversion cycles.
b. explain liquidity and compare issuers' liquidity levels.
c. describe issuers' objectives and compare methods for managing working capital and liquidity.

24. Capital Investments and Capital Allocation

The candidate should be able to:

a. describe types of capital investments.
b. describe the capital allocation process, calculate net present value (NPV), internal rate of return (IRR), and return on invested capital (ROIC), and contrast their use in capital allocation.
c. describe principles of capital allocation and common capital allocation pitfalls.
d. describe types of real options relevant to capital investments.

25. Capital Structure

The candidate should be able to:

a. calculate and interpret the weighted-average cost of capital for a company.
b. explain factors affecting capital structure and the weighted-average cost of capital.
c. explain the Modigliani–Miller propositions regarding capital structure.
d. describe optimal and target capital structures.

26. Business Models

The candidate should be able to:

a. describe key features of business models.
b. describe various types of business models.

READING 1

RATES AND RETURNS

MODULE 1.1: INTEREST RATES AND RETURN MEASUREMENT

Video covering this content is available online.

LOS 1.a: Interpret interest rates as required rates of return, discount rates, or opportunity costs and explain an interest rate as the sum of a real risk-free rate and premiums that compensate investors for bearing distinct types of risk.

Interest rates measure the time value of money, although risk differences in financial securities lead to differences in their equilibrium interest rates. Equilibrium interest rates are the **required rate of return** for a particular investment, in the sense that the market rate of return is the return that investors and savers require to get them to willingly lend their funds. Interest rates are also referred to as **discount rates** and, in fact, the terms are often used interchangeably. If an individual can borrow funds at an interest rate of 10%, then that individual should discount payments to be made in the future at that rate to get their equivalent value in current dollars or other currencies. Finally, we can also view interest rates as the **opportunity cost** of current consumption. If the market rate of interest on 1-year securities is 5%, earning an additional 5% is the opportunity forgone when current consumption is chosen rather than saving (postponing consumption).

The **real risk-free rate** of interest is a theoretical rate on a single-period loan that contains no expectation of inflation and zero probability of default. What the real risk-free rate represents in economic terms is **time preference**, the degree to which current consumption is preferred to equal future consumption.

When we speak of a real rate of return, we are referring to an investor's increase in purchasing power (after adjusting for inflation). Because expected inflation in future periods is not zero, the rates we observe on U.S. Treasury bills (T-bills), for example, are essentially risk-free rates, but not real rates of return. T-bill rates are nominal risk-free rates because they contain an **inflation premium**. This is the relation:

$$(1 + \text{nominal risk-free rate}) = (1 + \text{real risk-free rate})(1 + \text{expected inflation rate})$$

Often, including in many parts of the CFA curriculum, this relation is approximated as follows:

$$\text{nominal risk-free rate} \approx \text{real risk-free rate} + \text{expected inflation rate}$$

Securities may have one or more types of risk, and each added risk increases the required rate of return. These types of risks are as follows:

- **Default risk.** This is the risk that a borrower will not make the promised payments in a timely manner.
- **Liquidity risk.** This is the risk of receiving less than fair value for an investment if it must be sold quickly for cash.
- **Maturity risk.** As we will see in the Fixed Income topic area, the prices of longer-term bonds are more volatile than those of shorter-term bonds. Longer-maturity bonds have more maturity risk than shorter-term bonds and require a maturity risk premium.

Each of these risk factors is associated with a risk premium that we add to the nominal risk-free rate to adjust for greater default risk, less liquidity, and longer maturity relative to a liquid, short-term, default risk-free rate such as that on T-bills. We can write the following:

$$\begin{aligned}\text{nominal rate of interest} = {} & \text{real risk-free rate} \\ & + \text{inflation premium} \\ & + \text{default risk premium} \\ & + \text{liquidity premium} \\ & + \text{maturity premium}\end{aligned}$$

LOS 1.b: Calculate and interpret different approaches to return measurement over time and describe their appropriate uses.

Holding period return (HPR) is simply the percentage increase in the value of an investment over a given period:

$$\text{holding period return} = \frac{\text{end-of-period value}}{\text{beginning-of-period value}} - 1$$

For example, a stock that pays a dividend during a holding period has an HPR for that period equal to:

$$\frac{P_t + \text{Div}_t}{P_0} - 1, \text{ or } \frac{P_t - P_0 + \text{Div}_t}{P_0}$$

If a stock is valued at €20 at the beginning of the period, pays €1 in dividends over the period, and at the end of the period is valued at €22, the HPR is:

$$\text{HPR} = (22 + 1) / 20 - 1 = 0.15 = 15\%$$

Returns over multiple periods reflect compounding. For example, given HPRs for Years 1, 2, and 3, the HPR for the entire three-year period is:

$$\text{HPR} = (1 + \text{HPR}_{\text{Year 1}})(1 + \text{HPR}_{\text{Year 2}})(1 + \text{HPR}_{\text{Year 3}}) - 1$$

Later in this reading, we will see that a return over multiple years is typically stated as an *annualized return* rather than an HPR.

Average Returns

The **arithmetic mean return** is the simple average of a series of periodic returns. It has the statistical property of being an unbiased estimator of the true mean of the underlying distribution of returns:

$$\text{arithmetic mean return} = \frac{(R_1 + R_2 + R_3 + \ldots + R_n)}{n}$$

The **geometric mean return** is a compound rate. When periodic rates of return vary from period to period, the geometric mean return will have a value less than the arithmetic mean return:

$$\text{geometric mean return} = \sqrt[n]{(1 + R_1) \times (1 + R_2) \times (1 + R_3) \times \ldots \times (1 + R_n)} - 1$$

For example, for returns R_t over three annual periods, the geometric mean return is calculated as the following example shows.

EXAMPLE: Geometric mean return

For the last three years, the returns for Acme Corporation common stock have been −9.34%, 23.45%, and 8.92%. Calculate the compound annual rate of return over the three-year period.

Answer:

$$R_G = \sqrt[3]{(1 - 0.0934) \times (1 + 0.2345) \times (1 + 0.0892)} - 1$$
$$= \sqrt[3]{0.9066 \times 1.2345 \times 1.0892} - 1$$
$$= \sqrt[3]{1.21903} - 1$$
$$R_G = 1.06825 - 1 = 6.825\%$$

Solve this type of problem with your calculator as follows:

- On the TI, enter 1.21903 [y^x] 3 [1/x] [=]
- On the HP, enter 1.21903 [ENTER] 3 [1/x] [y^x]

In the previous example, the geometric mean results in an annual rate of return because the holding periods were years. If the holding periods are other than years, the geometric mean is not the same as the annual return. The root for the geometric mean is the number of *periods*, while the root for the annual return is the number of *years*.

EXAMPLE: Geometric mean vs. annual return

For the last four semiannual periods, the 6-month holding period returns on an investment were 2.0%, 0.5%, −1.0%, and 1.5%. Calculate the geometric mean and the annual rate of return.

Answer:

$$\text{Geometric mean} = \sqrt[4]{(1+0.02)(1+0.005)(1-0.10)(1+0.015)} - 1$$
$$= 0.007435 = 0.7435\%.$$

This is the geometric mean of the 6-month holding period returns.

$$\text{Annual return} = \sqrt[2]{(1+0.02)(1+0.005)(1-0.10)(1+0.015)} - 1$$
$$= 0.0149 = 1.49\%.$$

The four semiannual periods equal two years, so to get an annual return we use 2 as the root.



PROFESSOR'S NOTE

The geometric mean is always less than or equal to the arithmetic mean, and the difference increases as the dispersion of the observations increases. The only time the arithmetic and geometric means are equal is when there is no variability in the observations (i.e., all observations are equal).

A **harmonic mean** is used for certain computations, such as the average cost of shares purchased over time. The harmonic mean is calculated as $\frac{N}{\sum_{i=1}^{N}\frac{1}{X_i}}$ where there are *N* values of X_i.

EXAMPLE: Calculating average cost with the harmonic mean

An investor purchases $1,000 of mutual fund shares each month, and over the last three months, the prices paid per share were $8, $9, and $10. What is the average cost per share?

Answer:

$$\overline{X}_H = \frac{3}{\frac{1}{8}+\frac{1}{9}+\frac{1}{10}} = \$8.926 \text{ per share}$$

To check this result, calculate the total shares purchased as follows:

$$\frac{1{,}000}{8}+\frac{1{,}000}{9}+\frac{1{,}000}{10} = 336.11 \text{ shares}$$

$$\text{The average price is } \frac{\$3{,}000}{336.11} = \$8.926 \text{ per share.}$$

The previous example illustrates the interpretation of the harmonic mean in its most common application. Note that the average price paid per share ($8.93) is less than the arithmetic average of the share prices, which is $\frac{8+9+10}{3} = 9$.

We can only calculate a harmonic mean of positive numbers. For a set of returns that includes negative numbers, we can treat them the same way we did with geometric means, using (1 + return) for each period, then subtracting 1 from the result.

EXAMPLE: Harmonic mean with negative returns

For four periods, the returns on an investment were 2.0%, 0.5%, −1.0%, and 1.5%. Calculate the harmonic mean of these returns.

Answer:

$$\text{Harmonic mean} = \frac{4}{\frac{1}{(1+0.02)} + \frac{1}{(1+0.005)} + \frac{1}{(1-0.01)} + \frac{1}{(1+0.015)}} - 1$$

$$= 0.007369 = 0.7369\%$$

The relationship among arithmetic, geometric, and harmonic means can be stated as follows:

$$\text{arithmetic mean} \times \text{harmonic mean} = (\text{geometric mean})^2$$

PROFESSOR'S NOTE
The proof of this is beyond the scope of the Level I exam.

For values that are not all equal, harmonic mean < geometric mean < arithmetic mean. This mathematical fact is the basis for the claimed benefit of purchasing the same money amount of mutual fund shares each month or each week. Some refer to this practice as **cost averaging**.

Measures of average return can be affected by outliers, which are unusual observations in a dataset. Two of the methods for dealing with outliers are a *trimmed mean* and a *winsorized mean*. We will examine these in our reading on Statistical Measures of Asset Returns.

Appropriate uses for the various return measures are as follows:

- **Arithmetic mean.** Include all values, including outliers.
- **Geometric mean.** Compound the rate of returns over multiple periods.
- **Harmonic mean.** Calculate the average share cost from periodic purchases in a fixed money amount.
- **Trimmed or winsorized mean.** Decrease the effect of outliers.

MODULE QUIZ 1.1

1. An interest rate is *best* interpreted as a:
   A. discount rate or a measure of risk.
   B. measure of risk or a required rate of return.
   C. required rate of return or the opportunity cost of consumption.

2. An interest rate from which the inflation premium has been subtracted is known as a:
   A. real interest rate.
   B. risk-free interest rate.
   C. real risk-free interest rate.

3. The harmonic mean of 3, 4, and 5 is:
   A. 3.74.
   B. 3.83.

C. 4.12.

4. **XYZ Corp. Annual Stock Returns**

Year	20X1	20X2	20X3	20X4	20X5	20X6
Return	22%	5%	−7%	11%	2%	11%

The mean annual return on XYZ stock is *most appropriately* calculated using the:

A. harmonic mean.
B. arithmetic mean.
C. geometric mean.

MODULE 1.2: TIME-WEIGHTED AND MONEY-WEIGHTED RETURNS

Video covering this content is available online.

LOS 1.c: Compare the money-weighted and time-weighted rates of return and evaluate the performance of portfolios based on these measures.

The **money-weighted return** applies the concept of the **internal rate of return (IRR)** to investment portfolios. An IRR is the interest rate at which a series of cash inflows and outflows sum to zero when discounted to their present value. That is, they have a **net present value (NPV)** of zero. The IRR and NPV are built-in functions on financial calculators that CFA Institute permits candidates to use for the exam.

PROFESSOR'S NOTE
We have provided an online video in the Resource Library on how to use the TI calculator. You can view it by logging in to your account at www.schweser.com.

The **money-weighted rate of return** is defined as the IRR on a portfolio, taking into account all cash inflows and outflows. The beginning value of the account is an inflow, as are all deposits into the account. All withdrawals from the account are outflows, as is the ending value.

EXAMPLE: Money-weighted rate of return

Assume an investor buys a share of stock for \$100 at $t = 0$, and at the end of the year ($t = 1$), she buys an additional share for \$120. At the end of Year 2, the investor sells both shares for \$130 each. At the end of each year in the holding period, the stock paid a \$2 per share dividend. What is the money-weighted rate of return?

Step 1: Determine the timing of each cash flow and whether the cash flow is an inflow (+), into the account, or an outflow (−), available from the account.

$t = 0$:	purchase of first share	= +$100.00	inflow to account
$t = 1$:	purchase of second share	= +$120.00	
	dividends from first share	= −$2.00	
	subtotal, $t = 1$	+$118.00	inflow to account
$t = 2$:	dividend from two shares	= −$4.00	
	proceeds from selling shares	= −$260.00	
	subtotal, $t = 2$	−$264.00	outflow from account

Step 2: Net the cash flows for each period and set the PV of cash inflows equal to the PV of cash outflows.

$$PV_{inflows} = PV_{outflows}$$

$$\$100 + \frac{\$118}{(1+r)} = \frac{\$264}{(1+r)^2}$$

Step 3: Solve for r to find the money-weighted rate of return. This can be done using trial and error or by using the IRR function on a financial calculator or spreadsheet.

The intuition here is that we deposited $100 into the account at t = 0, then added $118 to the account at t = 1 (which, with the $2 dividend, funded the purchase of one more share at $120), and ended with a total value of $264.

To compute this value with a financial calculator, use these net cash flows and follow the procedure(s) described to calculate the IRR:

net cash flows: $CF_0 = +100$; $CF_1 = +120 - 2 = +118$;
$CF_2 = -260 + -4 = -264$

Calculating money-weighted return with the TI Business Analyst II Plus®

Note the values for F01, F02, and so on, are all equal to 1.

Keystrokes	Explanation	Display
[CF] [2nd][CLR WORK]	Clear cash flow registers	CF0 = 0.00000
100 [ENTER]	Initial cash outlay	CF0 = +100.00000
[↓] 118 [ENTER]	Period 1 cash flow	C01 = +118.00000
[↓] [↓] 264 [+/−] [ENTER]	Period 2 cash flow	C02 = −264.00000
[IRR] [CPT]	Calculate IRR	IRR = 13.86122

The money-weighted rate of return for this problem is 13.86%.

PROFESSOR'S NOTE

In the preceding example, we entered the flows into the account as a positive and the ending value as a negative (the investor could withdraw this amount from the account). Note that there is no difference in the solution if we enter the cash flows into the account as negative values (out of the investor's

pocket) and the ending value as a positive value (into the investor's pocket). As long as payments into the account and payments out of the account (including the ending value) are entered with opposite signs, the computed IRR will be correct.

Time-weighted rate of return measures compound growth and is the rate at which $1 compounds over a specified performance horizon. Time-weighting is the process of averaging a set of values over time. The annual time-weighted return for an investment may be computed by performing the following steps:

Step 1: Value the portfolio immediately preceding significant additions or withdrawals. Form subperiods over the evaluation period that correspond to the dates of deposits and withdrawals.

Step 2: Compute the holding period return (HPR) of the portfolio for each subperiod.

Step 3: Compute the product of (1 + HPR) for each subperiod to obtain a total return for the entire measurement period [i.e., $(1 + HPR_1) \times (1 + HPR_2) \ldots (1 + HPR_n)] - 1$. If the total investment period is greater than one year, you must take the geometric mean of the measurement period return to find the annual time-weighted rate of return.

EXAMPLE: Time-weighted rate of return

An investor purchases a share of stock at $t = 0$ for $100. At the end of the year, $t = 1$, the investor buys another share of the same stock for $120. At the end of Year 2, the investor sells both shares for $130 each. At the end of both Years 1 and 2, the stock paid a $2 per share dividend. What is the annual time-weighted rate of return for this investment? (This is the same investment as the preceding example.)

Answer:

Step 1: Break the evaluation period into two subperiods based on timing of cash flows.

Holding period 1:	Beginning value	= $100
	Dividends paid	= $2
	Ending value	= $120
Holding period 2:	Beginning value	= $240 (2 shares)
	Dividends paid	= $4 ($2 per share)
	Ending value	= $260 (2 shares)

Step 2: Calculate the HPR for each holding period.

$$HPR_1 = [(\$120 + 2)/\$100] - 1 = 22\%$$

$$HPR_2 = [(\$260 + 4)/\$240] - 1 = 10\%$$

Step 3: Find the compound annual rate that would have produced a total return equal to the return on the account over the two-year period.

$$(1 + \text{time-weighted rate of return})^2 = (1.22)(1.10)$$

$$\text{time-weighted rate of return} = [(1.22)(1.10)]^{0.5} - 1 = 15.84\%$$

The time-weighted rate of return is not affected by the timing of cash inflows and outflows. In the investment management industry, time-weighted return is the preferred method of performance measurement because portfolio managers typically do not control the timing of deposits to and withdrawals from the accounts they manage.

In the preceding examples, the time-weighted rate of return for the portfolio was 15.84%, while the money-weighted rate of return for the same portfolio was 13.86%. The results are different because the money-weighted rate of return gave a larger weight to the Year 2 HPR, which was 10%, versus the 22% HPR for Year 1. This is because there was more money in the account at the beginning of the second period.

If funds are contributed to an investment portfolio just before a period of relatively poor portfolio performance, the money-weighted rate of return will tend to be lower than the time-weighted rate of return. On the other hand, if funds are contributed to a portfolio at a favorable time (just before a period of relatively high returns), the money-weighted rate of return will be higher than the time-weighted rate of return. The use of the time-weighted return removes these distortions, and thus provides a better measure of a manager's ability to select investments over the period. If the manager has complete control over money flows into and out of an account, the money-weighted rate of return would be the more appropriate performance measure.



MODULE QUIZ 1.2

1. An investor buys a share of stock for $40 at time $t = 0$, buys another share of the same stock for $50 at $t = 1$, and sells both shares for $60 each at $t = 2$. The stock paid a dividend of $1 per share at $t = 1$ and at $t = 2$. The periodic money-weighted rate of return on the investment is *closest* to:
   A. 22.2%. a
   B. 23.0%.
   C. 23.8%.

2. An investor buys a share of stock for $40 at time $t = 0$, buys another share of the same stock for $50 at $t = 1$, and sells both shares for $60 each at $t = 2$. The stock paid a dividend of $1 per share at $t = 1$ and at $t = 2$. The time-weighted rate of return on the investment for the period is *closest* to:
   A. 24.7%.
   B. 25.7%.
   C. 26.8%.

MODULE 1.3: COMMON MEASURES OF RETURN



Video covering this content is available online.

LOS 1.d: Calculate and interpret annualized return measures and continuously compounded returns, and describe their appropriate uses.

Interest rates and market returns are typically stated as **annualized returns**, regardless of the actual length of the time period over which they occur. To annualize an HPR that is realized over a specific number of days, use the following formula:

$$\text{annualized return} = (1 + \text{HPR})^{365/\text{days}} - 1$$

EXAMPLE: Annualized return, shorter than one year

A saver deposits $100 into a bank account. After 90 days, the account balance is $100.75. What is the saver's annualized rate of return?

Answer:

$$\text{HPR} = \frac{100.75}{100} - 1 = 0.0075 = 0.75\%$$
$$\text{annualized return} = (1 + 0.0075)^{365/90} - 1 = 0.0308 = 3.08\%$$

EXAMPLE: Annualized return, longer than one year

An investor buys a 500-day government bill for $970 and redeems it at maturity for $1,000. What is the investor's annualized return?

Answer:

$$\text{HPR} = \frac{1000}{970} - 1 = 0.0309 = 3.09\%$$
$$\text{annualized return} = (1 + 0.0309)^{365/500} - 1 = 0.0225 = 2.25\%$$

In time value of money calculations (which we will address in more detail in our reading on The Time Value of Money in Finance), more frequent compounding has an impact on future value and present value computations. Specifically, because an increase in the frequency of compounding increases the effective interest rate, it also *increases* the future value of a given cash flow and *decreases* the present value of a given cash flow.

This is the general formula for the present value of a future cash flow:

$$\text{PV} = \text{FV}_N\left(1 + \frac{r}{m}\right)^{-mN}$$

where:
r = quoted annual interest rate
N = number of years
m = compounding periods per year

EXAMPLE: The effect of compounding frequency on FV and PV

Compute the PV of $1,000 to be received one year from now using a stated annual interest rate of 6% with a range of compounding periods.

Answer:

With semiannual compounding, *m* = 2:

$$\text{PV} = 1{,}000\left(1 + \frac{0.06}{2}\right)^{-2} = 942.60$$

With quarterly compounding, *m* = 4:

$$PV = 1{,}000\left(1 + \frac{0.06}{4}\right)^{-4} = 942.18$$

With monthly compounding, $m = 12$:

$$PV = 1{,}000\left(1 + \frac{0.06}{12}\right)^{-12} = 941.91$$

With daily compounding, $m = 365$:

$$PV = 1{,}000\left(1 + \frac{0.06}{365}\right)^{-365} = 941.77$$

Compounding Frequency Effect

Compounding Frequency	Interest Rate per Period	Present Value
Annual ($m = 1$)	6.000%	$943.40
Semiannual ($m = 2$)	3.000	942.60
Quarterly ($m = 4$)	1.500	942.18
Monthly ($m = 12$)	0.500	941.91
Daily ($m = 365$)	0.016438	941.77

The mathematical limit of shortening the compounding period is known as continuous compounding. Given an HPR, we can use the natural logarithm (ln, or LN on your financial calculator) to state its associated **continuously compounded return**:

$$R_{CC} = \ln(1 + HPR) = \ln\left(\frac{\text{ending value}}{\text{beginning value}}\right)$$

Notice that because the calculation is based on 1 plus the HPR, we can also perform it directly from the **price relative**. The price relative is just the end-of-period value divided by the beginning-of-period value.

EXAMPLE: Calculating continuously compounded returns

A stock was purchased for $100 and sold one year later for $120. Calculate the investor's annual rate of return on a continuously compounded basis.

Answer:

$$\ln\left(\frac{120}{100}\right) = 18.232\%$$

If we had been given the return (20%) instead, the calculation is this:

$$\ln(1 + 0.20) = 18.232\%$$

A useful property of continuously compounded rates of return is that they are additive for multiple periods. That is, a continuously compounded return from $t = 0$ to $t = 2$ is

the sum of the continuously compounded return from $t = 0$ to $t = 1$ and the continuously compounded return from $t = 1$ to $t = 2$.

LOS 1.e: Calculate and interpret major return measures and describe their appropriate uses.

Gross return refers to the total return on a security portfolio before deducting fees for the management and administration of the investment account. **Net return** refers to the return after these fees have been deducted. Commissions on trades and other costs that are necessary to generate the investment returns are deducted in both gross and net return measures.

Pretax nominal return refers to the return before paying taxes. Dividend income, interest income, short-term capital gains, and long-term capital gains may all be taxed at different rates. **After-tax nominal return** refers to the return after the tax liability is deducted.

Real return is nominal return adjusted for inflation. Consider an investor who earns a nominal return of 7% over a year when inflation is 2%. The investor's approximate real return is simply 7 − 2 = 5%. The investor's exact real return is slightly lower: 1.07 / 1.02 − 1 = 0.049 = 4.9%.

Using the components of an interest rate we defined earlier, we can state a real return as follows:

$$(1 + \text{real return}) = \frac{(1 + \text{nominal risk-free rate})(1 + \text{risk premium})}{(1 + \text{inflation premium})}$$

PROFESSOR'S NOTE

The Level I curriculum states this relationship as

$$(1 + \text{real return}) = \frac{(1 + \text{real risk-free rate})(1 + \text{risk premium})}{(1 + \text{inflation premium})}.$$

Stating it this way assumes the risk premium includes inflation risk.

Real return measures the increase in an investor's purchasing power—how much more goods she can purchase at the end of one year due to the increase in the value of her investments. If she invests $1,000 and earns a nominal return of 7%, she will have $1,070 at the end of the year. If the price of the goods she consumes has gone up 2%, from $1.00 to $1.02, she will be able to consume 1,070 / 1.02 = 1,049 units. She has given up consuming 1,000 units today, but instead, she is able to purchase 1,049 units at the end of one year. Her purchasing power has gone up 4.9%; this is her real return.

A **leveraged return** refers to a return to an investor that is a multiple of the return on the underlying asset. The leveraged return is calculated as the gain or loss on the investment as a percentage of an investor's cash investment. An investment in a derivative security, such as a futures contract, produces a leveraged return because the cash deposited is only a fraction of the value of the assets underlying the futures contract. Leveraged investments in real estate are common: investors pay only a portion of a property's cost in cash and borrow the rest.

To illustrate the effect of leverage on returns, consider a fund that can invest the amount V_0 without leverage, and earn the rate of return *r*. The fund's unleveraged return (as a money amount) is simply $r \times V_0$. Now let's say this fund can borrow the amount V_B at an interest rate of r_B, and earn *r* by investing the proceeds. The fund's leveraged return (again as a money amount), after subtracting the interest cost, then becomes $r \times (V_0 + V_B) - (r_B \times V_B)$.

Thus, stated as a rate of return on the initial value of V_0, the leveraged rate of return is as follows:

$$\text{leveraged return} = \frac{r(V_0 + V_B) - r_B V_B}{V_0}$$



MODULE QUIZ 1.3

1. If an investment loses 3% of its value over 120 days, its annualized return is *closest* to:
   A. −8.0%.
   B. −8.5%.
   C. −9.0%.
2. If a stock's initial price is $20 and its price increases to $23, its continuously compounded rate of return is *closest* to:
   A. 13.64%.
   B. 13.98%.
   C. 15.00%.
3. The value of an investment increases 5% before commissions and fees. This 5% increase represents:
   A. the investment's net return.
   B. the investment's gross return.
   C. neither the investment's gross return nor its net return.

KEY CONCEPTS

LOS 1.a

An interest rate can be interpreted as the rate of return required in equilibrium for a particular investment, the discount rate for calculating the present value of future cash flows, or as the opportunity cost of consuming now, rather than saving and investing.

The real risk-free rate reflects time preference for present goods versus future goods. Nominal risk-free rate ≈ real risk-free rate + expected inflation rate.

Securities may have several risks, and each increases the required rate of return. These include default risk, liquidity risk, and maturity risk.

We can view a nominal interest rate as the sum of a real risk-free rate, expected inflation, a default risk premium, a liquidity premium, and a maturity premium.

LOS 1.b

Holding period return is used to measure an investment's return over a specific period. Arithmetic mean return is the simple average of a series of periodic returns. Geometric mean return is a compound annual rate.

Arithmetic mean return includes all observations, including outliers. Geometric mean return is used for compound returns over multiple periods. Harmonic mean is used to calculate the average price paid with equal periodic investments. Trimmed mean or winsorized mean are used to reduce the effect of outliers.

LOS 1.c

The money-weighted rate of return is the IRR calculated using periodic cash flows into and out of an account and is the discount rate that makes the PV of cash inflows equal to the PV of cash outflows.

The time-weighted rate of return measures compound growth and is the rate at which money compounds over a specified performance horizon.

If funds are added to a portfolio just before a period of poor performance, the money-weighted return will be lower than the time-weighted return. If funds are added just before a period of high returns, the money-weighted return will be higher than the time-weighted return.

The time-weighted return is the preferred measure of a manager's ability to select investments. If the manager controls the money flows into and out of an account, the money-weighted return is the more appropriate performance measure.

LOS 1.d

Interest rates and market returns are typically stated on an annualized basis:

$$\text{annualized return} = (1 + \text{HPR})^{365/\text{days}} - 1$$

Given a holding period return, this is the associated continuously compounded return:

$$R_{cc} = \ln(1 + \text{HPR}) = \ln\left(\frac{\text{ending value}}{\text{beginning value}}\right)$$

LOS 1.e

Gross return is the total return after deducting commissions on trades and other costs necessary to generate the returns, but before deducting fees for the management and administration of the investment account. Net return is the return after management and administration fees have been deducted.

Pretax nominal return is the numerical percentage return of an investment, without considering the effects of taxes and inflation. After-tax nominal return is the numerical return after the tax liability is deducted, without adjusting for inflation.

Real return is the increase in an investor's purchasing power, roughly equal to nominal return minus inflation.

Leveraged return is the gain or loss on an investment as a percentage of an investor's cash investment.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 1.1

1. **C** Interest rates can be interpreted as required rates of return, discount rates, or opportunity costs of current consumption. A risk premium can be, but is not always, a component of an interest rate. (LOS 1.a)

2. **A** Real interest rates are those that have been adjusted for inflation. (LOS 1.a)

3. **B** $\overline{X}_H = \frac{3}{\frac{1}{3}+\frac{1}{4}+\frac{1}{5}} = 3.83$

   (LOS 1.b)

4. **C** Because returns are compounded, the geometric mean is appropriate.

   $[(1.22)(1.05)(0.93)(1.11)(1.02)(1.11)]^{1/6} - 1 = 6.96\%$

   (LOS 1.b)

Module Quiz 1.2

1. **C** Using the cash flow functions on your financial calculator, enter $CF_0 = -40$; $CF_1 = -50 + 1 = -49$; $CF_2 = 60 \times 2 + 2 = 122$; and CPT IRR = 23.82%. (LOS 1.c)

2. **A** $HPR_1 = \frac{50+1}{40} - 1 = 27.5\%$

   $HPR_2 = \frac{120+2}{100} - 1 = 22.0\%$

   $TWR = \sqrt{(1+0.275)(1+0.22)} - 1 = 24.72\%$

   (LOS 1.c)

Module Quiz 1.3

1. **C** $\text{Annualized return} = (1-0.03)^{365/120} - 1 = -0.0885 = -8.85\%$

   (LOS 1.d)

2. **B** ln (23/20) = 0.1398

   (LOS 1.d)

3. **C** Gross return is the total return after deducting commissions on trades and other costs necessary to generate the returns, but before deducting fees for the management and administration of the investment account. Net return is the return after management and administration fees have been deducted. (LOS 1.e)

READING 2

THE TIME VALUE OF MONEY IN FINANCE



PROFESSOR'S NOTE
The examples we use in this reading are meant to show how the time value of money appears throughout finance. Don't worry if you are not yet familiar with the securities we describe in this reading. We will see these examples again when we cover bonds and forward interest rates in Fixed Income, stocks in Equity Investments, foreign exchange in Economics, and options in Derivatives.

WARM-UP: USING A FINANCIAL CALCULATOR

For the exam, you must be able to use a financial calculator when working time value of money problems. You simply do not have the time to solve these problems any other way.

CFA Institute allows only two types of calculators to be used for the exam: (1) the Texas Instruments® TI BA II Plus™ (including the BA II Plus Professional™) and (2) the HP® 12C (including the HP 12C Platinum). This reading is written primarily with the TI BA II Plus in mind. If you do not already own a calculator, purchase a TI BA II Plus! However, if you already own the HP 12C and are comfortable with it, by all means, continue to use it.

Before we begin working with financial calculators, you should familiarize yourself with your TI BA II Plus by locating the keys noted below. These are the only keys you need to know to calculate virtually all of the time value of money problems:

- N = number of compounding periods
- I/Y = interest rate per compounding period
- PV = present value
- FV = future value
- PMT = annuity payments, or constant periodic cash flow
- CPT = compute

The TI BA II Plus comes preloaded from the factory with the periods per year function (P/Y) set to 12. This automatically converts the annual interest rate (I/Y) into monthly rates. While appropriate for many loan-type problems, this feature is not suitable for the vast majority of the time value of money applications we will be studying. So,

before using our SchweserNotes™, please set your P/Y key to "1" using the following sequence of keystrokes:

[2nd] [P/Y] "1" [ENTER] [2nd] [QUIT]

As long as you do not change the P/Y setting, it will remain set at one period per year until the battery from your calculator is removed (it does not change when you turn the calculator on and off). If you want to check this setting at any time, press [2nd] [P/Y]. The display should read P/Y = 1.0. If it does, press [2nd] [QUIT] to get out of the "programming" mode. If it does not, repeat the procedure previously described to set the P/Y key. With P/Y set to equal 1, it is now possible to think of I/Y as the interest rate per compounding period and *N* as the number of compounding periods under analysis. Thinking of these keys in this way should help you keep things straight as we work through time value of money problems.

PROFESSOR'S NOTE
We have provided an online video in the Resource Library on how to use the TI calculator. You can view it by logging in to your account at www.schweser.com.

MODULE 2.1: DISCOUNTED CASH FLOW VALUATION

Video covering this content is available online.

LOS 2.a: Calculate and interpret the present value (PV) of fixed-income and equity instruments based on expected future cash flows.

In our Rates and Returns reading, we gave examples of the relationship between present values and future values. We can simplify that relationship as follows:

$$FV = PV(1 + r)^t$$

$$PV = \frac{FV}{(1 + r)^t} = FV(1 + r)^{-t}$$

where:
r = interest rate per compounding period
t = number of compounding periods

If we are using continuous compounding, this is the relationship:

$$FV = PV \times e^{rt}$$

$$PV = FV \times e^{-rt}$$

Fixed-Income Securities

One of the simplest examples of the time value of money concept is a **pure discount** debt instrument, such as a **zero-coupon bond**. With a pure discount instrument, the investor pays less than the face value to buy the instrument and receives the face value at maturity. The price the investor pays depends on the instrument's **yield to maturity** (the discount rate applied to the face value) and the time until maturity. The amount of

interest the investor earns is the difference between the face value and the purchase price.

EXAMPLE: Zero-coupon bond

A zero-coupon bond with a face value of $1,000 will mature 15 years from today. The bond has a yield to maturity of 4%. Assuming annual compounding, what is the bond's price?

Answer:

$$\text{PV} = \frac{\$1{,}000}{(1 + 0.04)^{15}} = \$555.26$$

We can infer a bond's yield from its price using the same relationship. Rather than solving for *r* with algebra, we typically use our financial calculators. For this example, if we were given the price of $555.26, the face value of $1,000, and annual compounding over 15 years, we would enter the following:

$$\text{PV} = -555.26$$

$$\text{FV} = 1{,}000$$

$$\text{PMT} = 0$$

$$\text{N} = 15$$

Then, to get the yield, CPT I/Y = 4.00.



PROFESSOR'S NOTE
Remember to enter cash outflows as negative values and cash inflows as positive values. From the investor's point of view, the purchase price (PV) is an outflow, and the return of the face value at maturity (FV) is an inflow.

In some circumstances, interest rates can be negative. A zero-coupon bond with a negative yield would be priced at a **premium**, which means its price is greater than its face value.

EXAMPLE: Zero-coupon bond with a negative yield

If the bond in the previous example has a yield to maturity of −0.5%, what is its price, assuming annual compounding?

Answer:

$$\text{PV} = \frac{\$1{,}000}{(1 - 0.005)^{15}} = \$1{,}078.09$$

A **fixed-coupon bond** is only slightly more complex. With a coupon bond, the investor receives a cash interest payment each period in addition to the face value at maturity. The bond's **coupon rate** is a percentage of the face value and determines the amount of

the interest payments. For example, a 3% annual coupon, $1,000 bond pays 3% of $1,000, or $30, each year.

The coupon rate and the yield to maturity are two different things. We only use the coupon rate to determine the coupon payment (PMT). The yield to maturity (I/Y) is the discount rate implied by the bond's price.

EXAMPLE: Price of an annual coupon bond

Consider a 10-year, $1,000 par value, 10% coupon, annual-pay bond. What is the value of this bond if its yield to maturity is 8%?

Answer:

The coupon payments will be 10% × $1,000 = $100 at the end of each year. The $1,000 par value will be paid at the end of Year 10, along with the last coupon payment.

The value of this bond with a discount rate (yield to maturity) of 8% is:

$$\frac{100}{1.08}+\frac{100}{1.08^2}+\frac{100}{1.08^3}+\ldots+\frac{100}{1.08^9}+\frac{1{,}100}{1.08^{10}}=1{,}134.20$$

The calculator solution is:

$$\text{N}=10;\ \text{PMT}=100;\ \text{FV}=1{,}000;\ \text{I/Y}=8;\ \text{CPT PV}=-1{,}134.20$$

The bond's value is $1,134.20.

PROFESSOR'S NOTE
For this reading where we want to illustrate time value of money concepts, we are only using annual coupon payments and compounding periods. In the Fixed Income topic area, we will also perform these calculations for semiannual-pay bonds.

Some bonds exist that have no maturity date. We refer to these as **perpetual bonds** or **perpetuities**. We cannot speak meaningfully of the future value of a perpetuity, but its present value simplifies mathematically to the following:

$$\text{PV of a perpetuity}=\frac{\text{payment}}{\text{r}}$$

An **amortizing bond** is one that pays a level amount each period, including its maturity period. The difference between an amortizing bond and a fixed-coupon bond is that for an amortizing bond, each payment includes some portion of the principal. With a fixed-coupon bond, the entire principal is paid to the investor on the maturity date.

Amortizing bonds are an example of an **annuity** instrument. For an annuity, the payment each period is calculated as follows:

$$\text{annuity payment} = \frac{r \times PV}{1 - (1 + r)^{-t}}$$

where:

r = interest rate per period

t = number of periods

PV = present value (principal)

We can also determine an annuity payment using a financial calculator.

EXAMPLE: Computing a loan payment

Suppose you are considering applying for a $2,000 loan that will be repaid with equal end-of-year payments over the next 13 years. If the annual interest rate for the loan is 6%, how much are your payments?

Answer:

The size of the end-of-year loan payment can be determined by inputting values for the three known variables and computing PMT. Note that FV = 0 because the loan will be fully paid off after the last payment:

$$N = 13;\ I/Y = 6;\ PV = -2{,}000;\ FV = 0;\ CPT \rightarrow PMT = \$225.92$$

Equity Securities

As with fixed-income securities, we value **equity securities** such as common and preferred stock as the present value of their future cash flows. The key differences are that equity securities do not mature, and their cash flows may change over time.

Preferred stock pays a fixed dividend that is stated as a percentage of its **par value** (similar to the face value of a bond). As with bonds, we must distinguish between the stated percentage that determines the cash flows and the discount rate we apply to the cash flows. We say that equity investors have a **required return** that will induce them to own an equity share. This required return is the discount rate we use to value equity securities.

Because we can consider a preferred stock's fixed stream of dividends to be infinite, we can use the perpetuity formula to determine its value:

$$\text{preferred stock value} = \frac{D_p}{k_p}$$

where:

D_p = dividend per period

k_p = the market's required return on the preferred stock

EXAMPLE: Preferred stock valuation

A company's $100 par preferred stock pays a $5.00 annual dividend and has a required return of 8%. Calculate the value of the preferred stock.

Answer:

Value of the preferred stock: $D_p/k_P = \$5.00/0.08 = \$62.50$

Common stock is a residual claim to a company's assets after it satisfies all other claims. Common stock typically does not promise a fixed dividend payment. Instead, the company's management decides whether and when to pay common dividends.

Because the future cash flows are uncertain, we must use models to estimate the value of common stock. Here, we will look at three approaches analysts use frequently, which we call **dividend discount models (DDMs)**. We will return to these examples in the Equity Investments topic area and explain when each model is appropriate.

1. *Assume a constant future dividend*. Under this assumption, we can value a common stock the same way we value a preferred stock, using the perpetuity formula.
2. *Assume a constant growth rate of dividends*. With this assumption, we can apply the **constant growth DDM**, also known as the **Gordon growth model**. In this model, we state the value of a common share as follows:

$$V_0 = \frac{D_1}{k_e - g_c}$$

where:

V_0 = value of a share *this* period

D_1 = dividend expected to be paid *next* period

k_e = required return on common equity

g_c = constant growth rate of dividends

In this model, V_0 represents the PV of all the dividends in future periods, beginning with D_1. Note that k_e must be greater than g_c or the math will not work.

EXAMPLE: Gordon growth model valuation

Calculate the value of a stock that is expected to pay a $1.62 dividend next year, if dividends are expected to grow at 8% forever and the required return on equity is 12%.

Answer:

$$\text{Calculate the stock's value} = D_1/(k_e - g_c)$$
$$= \$1.62/(0.12 - 0.08)$$
$$= \$40.50$$

3. *Assume a changing growth rate of dividends*. This can be done in many ways. The example we will use here (and the one that is required for the Level I CFA exam) is known as a **multistage DDM**. Essentially, we assume a pattern of dividends in the short term, such as a period of high growth, followed by a constant growth rate of dividends in the long term.

   To use a multistage DDM, we discount the expected dividends in the short term as individual cash flows, then apply the constant growth DDM to the long term. As we saw in the previous example, the constant growth DDM gives us a value for an equity share *one period before* the dividend we use in the numerator.

EXAMPLE: Multistage growth

Consider a stock with dividends that are expected to grow at 15% per year for two years, after which they are expected to grow at 5% per year, indefinitely. The last dividend paid was \$1.00, and k_e = 11%. Calculate the value of this stock using the multistage growth model.

Answer:

Calculate the dividends over the high growth period:

$$D_1 = D_0(1 + g^*) = 1.00(1.15) = \$1.15$$
$$D_2 = D_1(1 + g^*) = 1.15(1.15) = 1.15^2 = \$1.32$$

Calculate the first dividend of the constant-growth period:

$$D_3 = D_2(1 + g) = 1.32 \times 1.05 = \$1.386$$

Use the constant growth model to get P_2, a value for all the (infinite) dividends expected from time = 3 onward:

$$P_2 = \frac{D_3}{k_e - g_c} = \frac{1.386}{0.11 - 0.05} = \$23.10$$

Finally, we can sum the present values of dividends 1 and 2 and of P_2 to get the present value of all the expected future dividends during both the high-growth and constant-growth periods:

$$\frac{1.15}{1.11} + \frac{1.32 + 23.10}{(1.11)^2} = \$20.86$$

PROFESSOR'S NOTE

A key point to notice in this example is that when we applied the dividend in Period 3 to the constant growth model, it gave us a value for the stock in Period 2. To get a value for the stock today, we had to discount this value back by two periods, along with the dividend in Period 2 that was not included in the constant growth value.



MODULE QUIZ 2.1

1. Terry Corporation preferred stock is expected to pay a $9 annual dividend in perpetuity. If the required rate of return on an equivalent investment is 11%, one share of Terry preferred should be worth:
   A. $81.82.
   B. $99.00.
   C. $122.22.

2. Dover Company wants to issue a $10 million face value of 10-year bonds with an annual coupon rate of 5%. If the investors' required yield on Dover's bonds is 6%, the amount the company will receive when it issues these bonds (ignoring transactions costs) will be:
   A. less than $10 million.
   B. equal to $10 million.
   C. greater than $10 million.

MODULE 2.2: IMPLIED RETURNS AND CASH FLOW ADDITIVITY



Video covering this content is available online.

LOS 2.b: Calculate and interpret the implied return of fixed-income instruments and required return and implied growth of equity instruments given the present value (PV) and cash flows.

The examples we have seen so far illustrate the relationships among present value, future cash flows, and the required rate of return. We can easily rearrange these relationships and solve for the required rate of return, given a security's price and its future cash flows.

EXAMPLE: Rate of return for a pure discount bond

A zero-coupon bond with a face value of $1,000 will mature 15 years from today. The bond's price is $650. Assuming annual compounding, what is the investor's annualized return?

Answer:

$$\frac{\$1,000}{(1+r)^{15}} = \$650$$

$$(1+r)^{15} = \frac{\$1,000}{\$650} = 1.5385$$

$$r = 1.5385^{1/15} - 1 = 0.0291 = 2.91\%$$

EXAMPLE: Yield of an annual coupon bond

Consider the 10-year, $1,000 par value, 10% coupon, annual-pay bond we examined in an earlier example, when its price was $1,134.20 at a yield to maturity of 8%. What is its yield to maturity if its price decreases to $1,085.00?

Answer:

$$\text{N} = 10;\ \text{PMT} = 100;\ \text{FV} = 1{,}000;\ \text{PV} = -1{,}085;\ \text{CPT I/Y} = 8.6934$$

The bond's yield to maturity increased to 8.69%.

Notice that the relationship between prices and yields is inverse. *When the price decreases, the yield to maturity increases. When the price increases, the yield to maturity decreases.* Or, equivalently, *when the yield increases, the price decreases. When the yield decreases, the price increases*. We will use this concept again and again when we study bonds in the Fixed Income topic area.

In our examples for equity share values, we assumed the investor's required rate of return. In practice, the required rate of return on equity is not directly observable. Instead, we use share prices that we can observe in the market to derive implied required rates of return on equity, given our assumptions about their future cash flows.

For example, if we assume a constant rate of dividend growth, we can rearrange the constant growth DDM to solve for the required rate of return:

$$V_0 = \frac{D_1}{k_e - g_c}$$

$$k_e - g_c = \frac{D_1}{V_0}$$

$$k_e = \frac{D_1}{V_0} + g_c$$

That is, the required rate of return on equity is the ratio of the expected dividend to the current price (which we refer to as a share's **dividend yield**) plus the assumed constant growth rate.

We can also rearrange the model to solve for a stock's **implied growth rate**, given a required rate of return:

$$k_e = \frac{D_1}{V_0} + g_c$$

$$g_c = k_e - \frac{D_1}{V_0}$$

That is, the implied growth rate is the required rate of return minus the dividend yield.

LOS 2.c: Explain the cash flow additivity principle, its importance for the no-arbitrage condition, and its use in calculating implied forward interest rates,

forward exchange rates, and option values.

The **cash flow additivity principle** refers to the fact that the PV of any stream of cash flows equals the sum of the PVs of the cash flows. If we have two series of cash flows, the sum of the PVs of the two series is the same as the PVs of the two series taken together, adding cash flows that will be paid at the same point in time. We can also divide up a series of cash flows any way we like, and the PV of the "pieces" will equal the PV of the original series.

EXAMPLE: Cash flow additivity principle

A security will make the following payments at the end of the next four years: $100, $100, $400, and $100. Calculate the PV of these cash flows using the concept of the PV of an annuity when the appropriate discount rate is 10%.

Answer:

We can divide the cash flows so that we have:

$t = 1$	$t = 2$	$t = 3$	$t = 4$	
100	100	100	100	Cash flow series #1
0	0	300	0	Cash flow series #2
$100	$100	$400	$100	

The additivity principle tells us that to get the PV of the original series, we can just add the PVs of cash flow series #1 (a 4-period annuity) and cash flow series #2 (a single payment three periods from now).

For the annuity: N = 4; PMT = 100; FV = 0; I/Y = 10; CPT →
PV = −$316.99

For the single payment: N = 3; PMT = 0; FV = 300; I/Y = 10; CPT →
PV = −$225.39

The sum of these two values is 316.99 + 225.39 = $542.38.

The sum of these two (present) values is identical (except for rounding) to the sum of the present values of the payments of the original series:

$$\frac{100}{1.1} + \frac{100}{1.1^2} + \frac{400}{1.1^3} + \frac{100}{1.1^4} = \$542.38$$

This is a simple example of **replication**. In effect, we created the equivalent of the given series of uneven cash flows by combining a 4-year annuity of 100 with a 3-year zero-coupon bond of 300.

We rely on the cash flow additivity principle in many of the pricing models we see in the Level I CFA curriculum. It is the basis for the **no-arbitrage principle**, or "law of one price," which says that if two sets of future cash flows are identical under all conditions, they will have the same price today (or if they don't, investors will quickly

buy the lower-priced one and sell the higher-priced one, which will drive their prices together).

Three examples of valuation based on the no-arbitrage condition are forward interest rates, forward exchange rates, and option pricing using a binomial model. We will explain each of these examples in greater detail when we address the related concepts in the Fixed Income, Economics, and Derivatives topic areas. For now, just focus on how they apply the principle that equivalent future cash flows must have the same present value.

Forward Interest Rates

A *forward interest rate* is the interest rate for a loan to be made at some future date. The notation used must identify both the length of the loan and when in the future the money will be borrowed. Thus, *1y1y* is the rate for a 1-year loan to be made one year from now; *2y1y* is the rate for a 1-year loan to be made two years from now; *3y2y* is the 2-year forward rate three years from now; and so on.

By contrast, a *spot interest rate* is an interest rate for a loan to be made today. We will use the notation S_1 for a 1-year rate today, S_2 for a 2-year rate today, and so on.

The way the cash flow additivity principle applies here is that, for example, borrowing for three years at the 3-year spot rate, or borrowing for one-year periods in three successive years, should have the same cost today. This relation is illustrated as follows: $(1 + S_3)^3 = (1 + S_1)(1 + 1y1y)(1 + 2y1y)$.

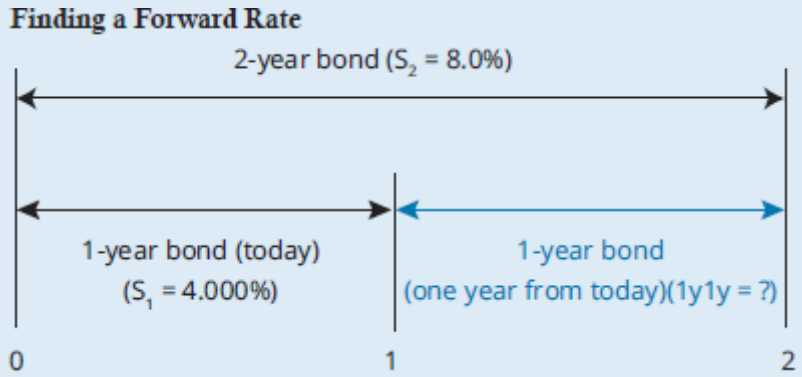
In fact, any combination of spot and forward interest rates that cover the same time period should have the same cost. Using this idea, we can derive **implied forward rates** from spot rates that are observable in the fixed-income markets.

EXAMPLE: Computing a forward rate from spot rates

The 2-period spot rate, S_2, is 8%, and the 1-period spot rate, S_1, is 4%. Calculate the forward rate for one period, one period from now, *1y1y*.

Answer:

The following figure illustrates the problem.



From our original equality, $(1 + S_2)^2 = (1 + S_1)(1 + 1y1y)$, we can get the following:

$$\frac{(1 + S_2)^2}{(1 + S_1)} = (1 + 1y1y)$$

Or, because we know that both choices have the same payoff in two years:

$$(1.08)^2 = (1.04)(1 + 1y1y)$$

$$(1 + 1y1y) = \frac{(1.08)^2}{(1.04)}$$

$$1y1y = \frac{(1.08)^2}{(1.04)} - 1 = \frac{1.1664}{1.04} - 1 = 12.154\%$$

In other words, investors are willing to accept 4.0% on the 1-year bond today (when they could get 8.0% on the 2-year bond today) only because they can get 12.154% on a 1-year bond one year from today. This future rate that can be locked in today is a forward rate.

Forward Currency Exchange Rates

An *exchange rate* is the price of one country's currency in terms of another country's currency. For example, an exchange rate of 1.416 USD/EUR means that one euro (EUR) is worth 1.416 U.S. dollars (USD). The Level I CFA curriculum refers to the currency in the numerator (USD, in this example) as the *price currency* and the one in the denominator (EUR in this example) as the *base currency*.

Like interest rates, exchange rates can be quoted as spot rates for currency exchanges to be made today, or as forward rates for currency exchanges to be made at a future date.

The percentage difference between forward and spot exchange rates is approximately the difference between the two countries' interest rates. This is because there is an arbitrage trade with a riskless profit to be made when this relation does not hold.

The possible arbitrage is as follows: borrow Currency A at Interest Rate A, convert it to Currency B at the spot rate and invest it to earn Interest Rate B, and sell the proceeds from this investment forward at the forward rate to turn it back into Currency A. If the forward rate does not correctly reflect the difference between interest rates, such an arbitrage could generate a profit to the extent that the return from investing Currency B and converting it back to Currency A with a forward contract is greater than the cost of borrowing Currency A for the period.

For spot and forward rates expressed as price currency/base currency, the no-arbitrage relation is as follows:

$$\frac{\text{forward}}{\text{spot}} = \frac{(1 + \text{interest rate}_{\text{price currency}})}{(1 + \text{interest rate}_{\text{base currency}})}$$

This formula can be rearranged as necessary to solve for specific values of the relevant terms.

EXAMPLE: Calculating the arbitrage-free forward exchange rate

Consider two currencies, the ABE and the DUB. The spot ABE/DUB exchange rate is 4.5671, the 1-year riskless ABE rate is 5%, and the 1-year riskless DUB rate is 3%. What is the 1-year forward exchange rate that will prevent arbitrage profits?

Answer:

Rearranging our formula, we have:

$$\text{forward} = \text{spot}\left(\frac{1 + I_{ABE}}{1 + I_{DUB}}\right)$$

and we can calculate the forward rate as:

$$\text{forward} = 4.5671\left(\frac{1.05}{1.03}\right) = 4.6558 \text{ ABE / DUB}$$

As you can see, the forward rate is greater than the spot rate by 4.6558 / 4.5671 − 1 = 1.94%. This is approximately equal to the interest rate differential of 5% − 3% = 2%.

Option Pricing Model

An *option* is the right, but not the obligation, to buy or sell an asset on a future date for a specified price. The right to buy an asset is a *call option*, and the right to sell an asset is a *put option*.

Valuing options is different from valuing other securities because the owner can let an option expire unexercised. A call option owner will let the option expire if the underlying asset can be bought in the market for less than the price specified in the option. A put option owner will let the option expire if the underlying asset can be sold in the market for more than the price specified in the option. In these cases, we say an option is *out of the money*. If an option is *in the money* on its expiration date, the owner has the right to buy the asset for less, or sell the asset for more, than its market price—and, therefore, will exercise the option.

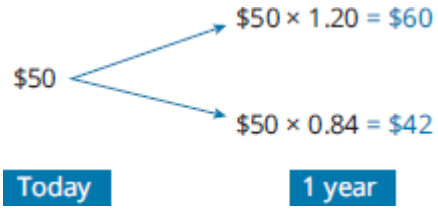
An approach to valuing options that we will use in the Derivatives topic area is a **binomial model**. A binomial model is based on the idea that, over the next period, some value will change to one of two possible values. To construct a one-period binomial model for pricing an option, we need the following:

- A value for the underlying asset at the beginning of the period
- An exercise price for the option; the exercise price can be different from the value of the underlying, and we assume the option expires one period from now
- Returns that will result from an up-move and a down-move in the value of the underlying over one period
- The risk-free rate over the period

As an example, we can model a call option with an exercise price of $55 on a stock that is currently valued (S_0) at $50. Let us assume that in one period, the stock's value will

either increase (S 1 u) to \$60 or decrease (S 1 d) to \$42. We state the return from an up-move (R^u) as \$60 / \$50 = 1.20, and the return from a down-move (R^d) as \$42 / \$50 = 0.84.

Figure 2.1: One-Period Binomial Tree



The call option will be in the money after an up-move or out of the money after a down-move. Its value at expiration after an up-move, c_1^u, is \$60 – \$55 = \$5. Its value after a down-move, c_1^d, is zero.

Now, we can use no-arbitrage pricing to determine the initial value of the call option (c_0). We do this by creating a portfolio of the option and the underlying stock, such that the portfolio will have the same value following either an up-move (V_1^u) or a down-move (V_1^d) in the stock. For our example, we would write the call option (that is, we grant someone else the option to buy the stock from us) and buy a number of shares of the stock that we will denote as *h*. We must solve for the *h* that results in $V_1^u = V_1^d$:

- The initial value of our portfolio, V_0, is $hS_0 - c_0$ (we subtract c_0 because we are short the call option).
- The portfolio value after an up-move, V_1^u, is $hS_1^u - c_1^u$.
- The portfolio value after a down-move, V_1^d, is $hS_1^d - c_1^d$.

In our example, $V_1^u = h(\$60) - \$5$, and $V_1^d = h(\$42) - 0$. Setting $V_1^u = V_1^d$ and solving for *h*, we get the following:

$$h(\$60) - \$5 = h(\$42)$$

$$h(\$60) - h(\$42) = \$5$$

$$h = \$5 / (\$60 - \$42) = 0.278$$

This result—the number of shares of the underlying we would buy for each call option we would write—is known as the hedge ratio for this option.

With $V_1^u = V_1^d$, the value of the portfolio after one period is known with certainty. This means we can say that either V_1^u or V_1^d must equal V_0 compounded at the risk-free rate for one period. In this example, V_1^d = 0.278(\$42) = \$11.68, or V_1^u = 0.278(\$60) – \$5 = \$11.68. Let us assume the risk-free rate over one period is 3%. Then, V_0 = \$11.68 / 1.03 = \$11.34.

Now, we can solve for the value of the call option, c_0. Recall that $V_0 = hS_0 - c_0$, so $c_0 = hS_0 - V_0$. Here, c_0 = 0.278(\$50) – \$11.34 = \$2.56.

MODULE QUIZ 2.2

1. For an equity share with a constant growth rate of dividends, we can estimate its:

A. value as the next dividend discounted at the required rate of return.
B. growth rate as the sum of its required rate of return and its dividend yield.
C. required return as the sum of its constant growth rate and its dividend yield.

2. An investment of €5 million today is expected to produce a one-time payoff of €7 million three years from today. The annual return on this investment, assuming annual compounding, is *closest* to:
A. 12%.
B. 13%.
C. 14%.

KEY CONCEPTS

LOS 2.a

The value of a fixed-income instrument or an equity security is the present value of its future cash flows, discounted at the investor's required rate of return:

$$PV = \frac{FV}{(1+r)^t} = FV(1+r)^{-t}$$

where:
r = interest rate per compounding period
t = number of compounding periods

$$\text{annuity payment} = \frac{r \times PV}{1-(1+r)^{-t}}$$

where:
r = interest rate per period
t = number of periods
PV = present value (principal)

The PV of a perpetual bond or a preferred stock $= \frac{\text{payment}}{r}$, where *r* = required rate of return.

The PV of a common stock with a constant growth rate of dividends is:

$$V_0 = \frac{D_1}{k_e - g_c}$$

LOS 2.b

By rearranging the present value relationship, we can calculate a security's required rate of return based on its price and its future cash flows. The relationship between prices and required rates of return is inverse.

For an equity share with a constant rate of dividend growth, we can estimate the required rate of return as the dividend yield plus the assumed constant growth rate, or we can estimate the implied growth rate as the required rate of return minus the dividend yield.

LOS 2.c

Using the cash flow additivity principle, we can divide up a series of cash flows any way we like, and the present value of the pieces will equal the present value of the original

series. This principle is the basis for the no-arbitrage condition, under which two sets of future cash flows that are identical must have the same present value.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 2.1

1. **A** 9 / 0.11 = $81.82 (LOS 2.a)

2. **A** Because the required yield is greater than the coupon rate, the present value of the bonds is less than their face value: N = 10; I/Y = 6; PMT = 0.05 × $10,000,00 = $500,000; FV = $10,000,000; and CPT PV = –$9,263,991. (LOS 2.a)

Module Quiz 2.2

1. **C** Using the constant growth dividend discount model, we can estimate the required rate of return as $k_e = \frac{D_1}{V_0} + g_c$. The estimated value of a share is *all* of its future dividends discounted at the required rate of return, which simplifies to $V_0 = \frac{D_1}{k_e - g_c}$ if we assume a constant growth rate. We can estimate the constant growth rate as the required rate of return *minus* the dividend yield. (LOS 2.b)

2. **A** $\left(\frac{7}{5}\right)^{\frac{1}{3}} - 1 = 0.1187$

   (LOS 2.b)

READING 3

STATISTICAL MEASURES OF ASSET RETURNS

MODULE 3.1: CENTRAL TENDENCY AND DISPERSION



Video covering this content is available online.

LOS 3.a: Calculate, interpret, and evaluate measures of central tendency and location to address an investment problem.

Measures of Central Tendency

Measures of central tendency identify the center, or average, of a dataset. This central point can then be used to represent the typical, or expected, value in the dataset.

The **arithmetic mean** is the sum of the observation values divided by the number of observations. It is the most widely used measure of central tendency. An example of an arithmetic mean is a **sample mean**, which is the sum of all the values in a sample of a population, ΣX, divided by the number of observations in the sample, *n*. It is used to make *inferences* about the population mean. The sample mean is expressed as follows:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

The **median** is the midpoint of a dataset, where the data are arranged in ascending or descending order. Half of the observations lie above the median, and half are below. To determine the median, arrange the data from the highest to lowest value, or lowest to highest value, and find the middle observation.

The median is important because the arithmetic mean can be affected by **outliers**, which are extremely large or small values. When this occurs, the median is a better measure of central tendency than the mean because it is not affected by extreme values that may actually be the result of errors in the data.

> **EXAMPLE: The median using an odd number of observations**
>
> What is the median return for five portfolio managers with a 10-year annualized total returns record of 30%, 15%, 25%, 21%, and 23%?

Answer:

First, arrange the returns in descending order:

30%, 25%, 23%, 21%, 15%

Then, select the observation that has an equal number of observations above and below it—the one in the middle. For the given dataset, the third observation, 23%, is the median value.

EXAMPLE: The median using an even number of observations

Suppose we add a sixth manager to the previous example with a return of 28%. What is the median return?

Answer:

Arranging the returns in descending order gives us this:

30%, 28%, 25%, 23%, 21%, 15%

With an even number of observations, there is no single middle value. The median value, in this case, is the arithmetic mean of the two middle observations, 25% and 23%. Thus, the median return for the six managers is 24% = 0.5(25 + 23).

The **mode** is the value that occurs most frequently in a dataset. A dataset may have more than one mode, or even no mode. When a distribution has one value that appears most frequently, it is said to be **unimodal**. When a dataset has two or three values that occur most frequently, it is said to be **bimodal** or **trimodal**, respectively.

EXAMPLE: The mode

What is the mode of the following dataset?

Dataset: [30%, 28%, 25%, 23%, 28%, 15%, 5%]

Answer:

The mode is 28% because it is the value appearing most frequently.

For continuous data, such as investment returns, we typically do not identify a single outcome as the mode. Instead, we divide the relevant range of outcomes into intervals, and we identify the **modal interval** as the one into which the largest number of observations fall.

Methods for Dealing With Outliers

In some cases, a researcher may decide that outliers should be excluded from a measure of central tendency. One technique for doing so is to use a **trimmed mean**. A trimmed mean excludes a stated percentage of the most extreme observations. A 1% trimmed mean, for example, would discard the lowest 0.5% and the highest 0.5% of the observations.

Another technique is to use a **winsorized mean**. Instead of discarding the highest and lowest observations, we substitute a value for them. To calculate a 90% winsorized mean, for example, we would determine the 5th and 95th percentile of the observations, substitute the 5th percentile for any values lower than that, substitute the 95th percentile for any values higher than that, and then calculate the mean of the revised dataset. Percentiles are measures of location, which we will address next.

Measures of Location

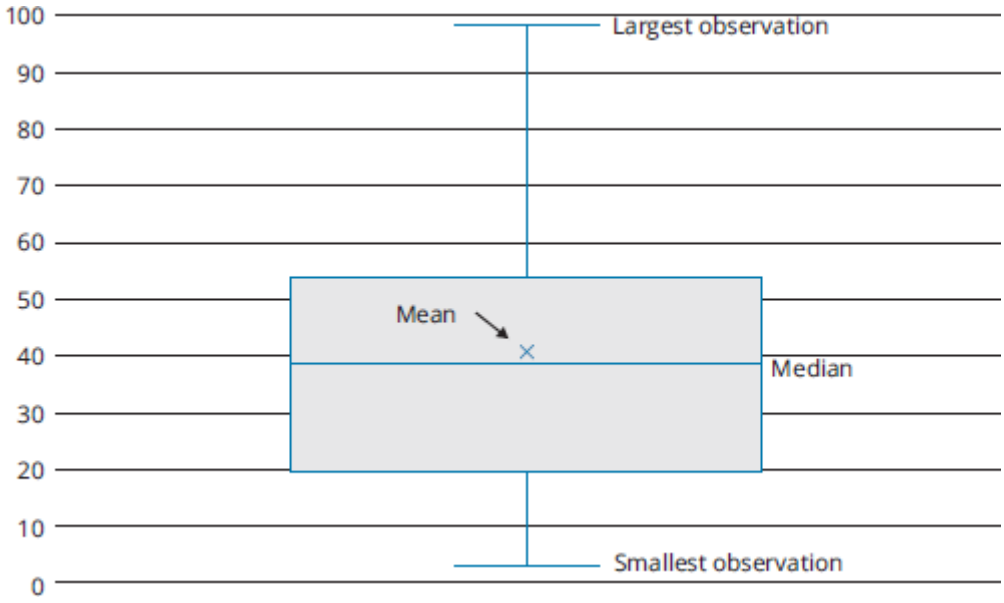
Quantile is the general term for a value at or below which a stated proportion of the data in a distribution lies. Examples of quantiles include the following:

- **Quartile.** The distribution is divided into quarters.
- **Quintile.** The distribution is divided into fifths.
- **Decile.** The distribution is divided into tenths.
- **Percentile.** The distribution is divided into hundredths (percentages).

Note that any quantile may be expressed as a percentile. For example, the third quartile partitions the distribution at a value such that three-fourths, or 75%, of the observations fall below that value. Thus, the third quartile is the 75th percentile. The difference between the third quartile and the first quartile (25th percentile) is known as the **interquartile range**.

To visualize a dataset based on quantiles, we can create a **box and whisker plot**, as shown in Figure 3.1. In a box and whisker plot, the box represents the central portion of the data, such as the interquartile range. The vertical line represents the entire range. In Figure 3.1, we can see that the largest observation is farther away from the center than is the smallest observation. This suggests that the data might include one or more outliers on the high side.

Figure 3.1: Box and Whisker Plot



LOS 3.b: Calculate, interpret, and evaluate measures of dispersion to address an investment problem.

Dispersion is defined as the *variability around the central tendency*. The common theme in finance and investments is the tradeoff between reward and variability, where the central tendency is the measure of the reward and dispersion is a measure of risk.

The **range** is a relatively simple measure of variability, but when used with other measures, it provides useful information. The range is the distance between the largest and the smallest value in the dataset:

range = maximum value − minimum value

> **EXAMPLE: The range**
>
> What is the range for the 5-year annualized total returns for five investment managers if the managers' individual returns were 30%, 12%, 25%, 20%, and 23%?
>
> **Answer:**
>
> range = 30 − 12 = 18%

The **mean absolute deviation (MAD)** is the average of the absolute values of the deviations of individual observations from the arithmetic mean:

$$\text{MAD} = \frac{\sum_{i=1}^{n} |X_i - \overline{X}|}{n}$$

The computation of the MAD uses the absolute values of each deviation from the mean because the sum of the actual deviations from the arithmetic mean is zero.

EXAMPLE: MAD

What is the MAD of the investment returns for the five managers discussed in the preceding example? How is it interpreted?

Answer:

Annualized returns: [30%, 12%, 25%, 20%, 23%]

$$\overline{X} = \frac{[30 + 12 + 25 + 20 + 23]}{5} = 22\%$$

$$MAD = \frac{[|30 - 22| + |12 - 22| + |25 - 22| + |20 - 22| + |23 - 22|]}{5}$$

$$MAD = \frac{[8 + 10 + 3 + 2 + 1]}{5} = 4.8\%$$

This result can be interpreted to mean that, on average, an individual return will deviate ±4.8% from the mean return of 22%.

The **sample variance**, s^2, is the measure of dispersion that applies when we are evaluating a sample of n observations from a population. The sample variance is calculated using the following formula:

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n - 1}$$

The denominator for s^2 is n − 1, one less than the sample size n. Based on the mathematical theory behind statistical procedures, the use of the entire number of sample observations, n, instead of n − 1 as the divisor in the computation of s^2, will systematically *underestimate* the population variance—particularly for small sample sizes. This systematic underestimation causes the sample variance to be a **biased estimator** of the population variance. Using n − 1 instead of n in the denominator, however, improves the statistical properties of s^2 as an estimator of the population variance.

EXAMPLE: Sample variance

Assume that the 5-year annualized total returns for the five investment managers used in the preceding examples represent only a sample of the managers at a large investment firm. What is the sample variance of these returns?

Answer:

$$\overline{X} = \frac{[30 + 12 + 25 + 20 + 23]}{5} = 22\%$$

$$s^2 = \frac{[(30 - 22)^2 + (12 - 22)^2 + (25 - 22)^2 + (20 - 22)^2 + (23 - 22)^2]}{5 - 1}$$

$$= 44.5(\%^2)$$

Thus, the sample variance of 44.5($\%^2$) can be interpreted to be an unbiased estimator of the population variance. Note that 44.5 "percent squared" is 0.00445, and you will get this value if you put the percentage returns in decimal form [e.g., $(0.30 - 0.22)^2$].

A major problem with using variance is the difficulty of interpreting it. The computed variance, unlike the mean, is in terms of squared units of measurement. How does one interpret squared percentages, squared dollars, or squared yen? This problem is mitigated through the use of the *standard deviation*. The units of standard deviation are the same as the units of the data (e.g., percentage return, dollars, euros). The **sample standard deviation** is the square root of the sample variance. The sample standard deviation, *s*, is calculated as follows:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}}$$

EXAMPLE: Sample standard deviation

Compute the sample standard deviation based on the result of the preceding example.

Answer:

Because the sample variance for the preceding example was computed to be 44.5($\%^2$), this is the sample standard deviation:

$$s = [44.5(\%^2)]^{1/2} = 6.67\%, \sqrt{0.00445} = 0.0667$$

This means that on average, an individual return from the sample will deviate ±6.67% from the mean return of 22%. The sample standard deviation can be interpreted as an unbiased estimator of the population standard deviation.

A direct comparison between two or more measures of dispersion may be difficult. For instance, suppose you are comparing the annual returns distribution for retail stocks with a mean of 8% and an annual returns distribution for a real estate portfolio with a mean of 16%. A direct comparison between the dispersion of the two distributions is not meaningful because of the relatively large difference in their means. To make a meaningful comparison, a relative measure of dispersion must be used. **Relative dispersion** is the amount of variability in a distribution around a reference point or benchmark. Relative dispersion is commonly measured with the **coefficient of variation (CV)**, which is computed as follows:

$$CV = \frac{s_x}{\overline{X}} = \frac{\text{standard deviation of x}}{\text{average value of x}}$$

CV measures the amount of dispersion in a distribution relative to the distribution's mean. This is useful because it enables us to compare dispersion across different sets of

data. In an investments setting, the CV is used to measure the risk (variability) per unit of expected return (mean). A lower CV is better.

EXAMPLE: Coefficient of variation

You have just been presented with a report that indicates that the mean monthly return on T-bills is 0.25% with a standard deviation of 0.36%, and the mean monthly return for the S&P 500 is 1.09% with a standard deviation of 7.30%. Your manager has asked you to compute the CV for these two investments and to interpret your results.

Answer:

$$CV_{\text{T-bills}} = \frac{0.36}{0.25} = 1.44$$

$$CV_{\text{S\&P 500}} = \frac{7.30}{1.09} = 6.70$$

These results indicate that there is less dispersion (risk) per unit of monthly return for T-bills than for the S&P 500 (1.44 vs. 6.70).



PROFESSOR'S NOTE
To remember the formula for CV, remember that the CV is a measure of variation, so standard deviation goes in the numerator. CV is variation per unit of return.

When we use variance or standard deviation as risk measures, we calculate risk based on outcomes both above and below the mean. In some situations, it may be more appropriate to consider only outcomes less than the mean (or some other specific value) in calculating a risk measure. In this case, we are measuring **downside risk**.

One measure of downside risk is **target downside deviation**, which is also known as **target semideviation**. Calculating target downside deviation is similar to calculating standard deviation, but in this case, we choose a target value against which to measure each outcome and only include deviations from the target value in our calculation if the outcomes are below that target.

The formula for target downside deviation is stated as follows:

$$s_{\text{target}} = \sqrt{\frac{\sum_{\text{all } X_i < B}^{n} (X_i - B)^2}{n - 1}}$$

where B = the target

Note that the denominator remains the sample size *n* minus one, even though we are not using all of the observations in the numerator.

EXAMPLE: Target downside deviation

Calculate the target downside deviation based on the data in the preceding examples, for a target return equal to the mean (22%), and for a target return of 24%.

Answer:

Return	Deviation From Mean	Deviation From Target Return
30%	30% − 22% = 8%	30% − 24% = 6%
12%	12% − 22% = −10%	12% − 24% = −12%
25%	25% − 22% = 3%	25% − 24% = 1%
20%	20% − 22% = −2%	20% − 24% = −4%
23%	23% − 22% = 1%	23% − 24% = −1%

$$s_{22\%} = \sqrt{\frac{(-10)^2 + (-2)^2}{5-1}} = 5.10\%$$

$$s_{24\%} = \sqrt{\frac{(-12)^2 + (-4)^2 + (-1)^2}{5-1}} = 6.34\%$$



MODULE QUIZ 3.1

1. A dataset has 100 observations. Which of the following measures of central tendency will be calculated using a denominator of 100?
   A. The winsorized mean, but not the trimmed mean.
   B. Both the trimmed mean and the winsorized mean.
   C. Neither the trimmed mean nor the winsorized mean.

2. **XYZ Corp. Annual Stock Returns**

20X1	20X2	20X3	20X4	20X5	20X6
22%	5%	−7%	11%	2%	11%

   What is the sample standard deviation?
   A. 9.8%.
   B. 72.4%.
   C. 96.3%.

3. **XYZ Corp. Annual Stock Returns**

20X1	20X2	20X3	20X4	20X5	20X6
22%	5%	−7%	11%	2%	11%

   Assume an investor has a target return of 11% for XYZ stock. What is the stock's target downside deviation?
   A. 9.39%.
   B. 12.10%.
   C. 14.80%.

MODULE 3.2: SKEWNESS, KURTOSIS, AND CORRELATION



Video covering this content is available online.

LOS 3.c: Interpret and evaluate measures of skewness and kurtosis to address an investment problem.

A distribution is symmetrical if it is shaped identically on both sides of its mean. Distributional symmetry implies that intervals of losses and gains will exhibit the same

frequency. For example, a symmetrical distribution with a mean return of zero will have losses in the −6% to −4% interval as frequently as it will have gains in the +4% to +6% interval. The extent to which a returns distribution is symmetrical is important because the degree of symmetry tells analysts if deviations from the mean are more likely to be positive or negative.

Skewness, or skew, refers to the extent to which a distribution is not symmetrical. Nonsymmetrical distributions may be either positively or negatively skewed and result from the occurrence of outliers in the dataset. **Outliers** are observations extraordinarily far from the mean, either above or below:

- A *positively skewed* distribution is characterized by outliers greater than the mean (in the upper region, or right tail). A positively skewed distribution is said to be skewed right because of its relatively long upper (right) tail.
- A *negatively skewed* distribution has a disproportionately large amount of outliers less than the mean that fall within its lower (left) tail. A negatively skewed distribution is said to be skewed left because of its long lower tail.

Skewness affects the location of the mean, median, and mode of a distribution:

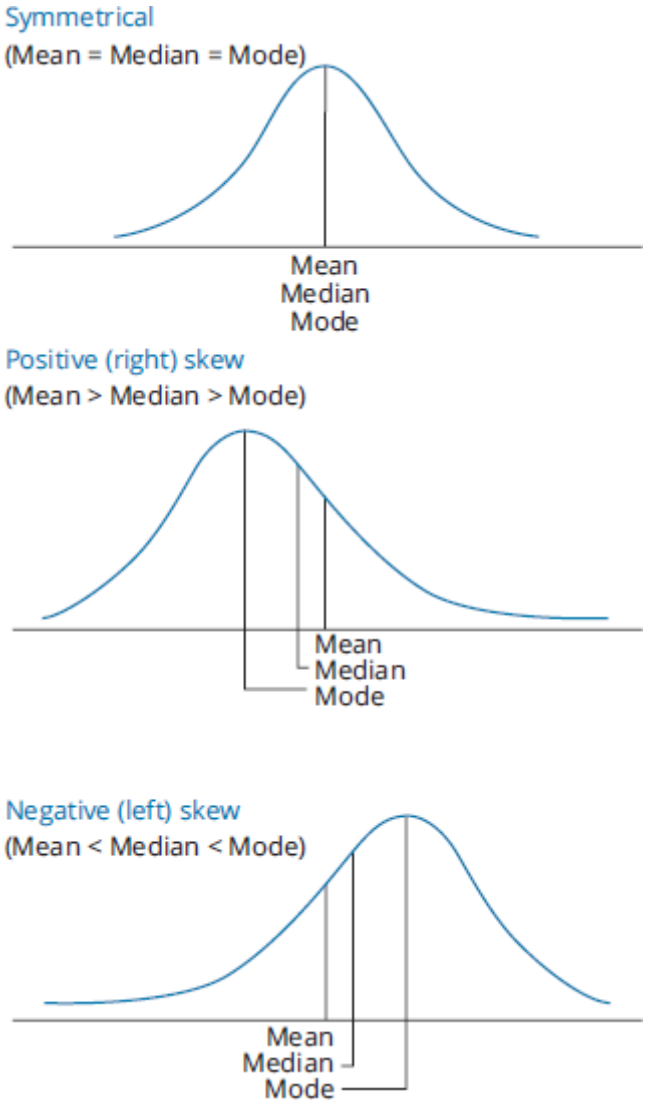
- For a symmetrical distribution, the mean, median, and mode are equal.
- For a positively skewed, unimodal distribution, the mode is less than the median, which is less than the mean. The mean is affected by outliers; in a positively skewed distribution, there are large, positive outliers, which will tend to pull the mean upward, or more positive. An example of a positively skewed distribution is that of housing prices. Suppose you live in a neighborhood with 100 homes; 99 of them sell for $100,000, and one sells for $1,000,000. The median and the mode will be $100,000, but the mean will be $109,000. Hence, the mean has been pulled upward (to the right) by the existence of one home (outlier) in the neighborhood.
- For a negatively skewed, unimodal distribution, the mean is less than the median, which is less than the mode. In this case, there are large, negative outliers that tend to pull the mean downward (to the left).



PROFESSOR'S NOTE

The key to remembering how measures of central tendency are affected by skewed data is to recognize that skew affects the mean more than the median and mode, and the mean is pulled in the direction of the skew. The relative location of the mean, median, and mode for different distribution shapes is shown in Figure 3.2. Note that the median is between the other two measures for positively or negatively skewed distributions.

Figure 3.2: Effect of Skewness on Mean, Median, and Mode



Sample skewness is equal to the sum of the cubed deviations from the mean divided by the cubed standard deviation and by the number of observations. Sample skewness for large samples is approximated as follows:

$$\text{sample skewness} \approx \left(\frac{1}{n}\right)\frac{\sum_{i=1}^{n}(X_i - \overline{X})^3}{s^3}$$

where:

s = sample standard deviation

PROFESSOR'S NOTE
The LOS requires us to "interpret and evaluate" measures of skewness and kurtosis, but not to calculate them.

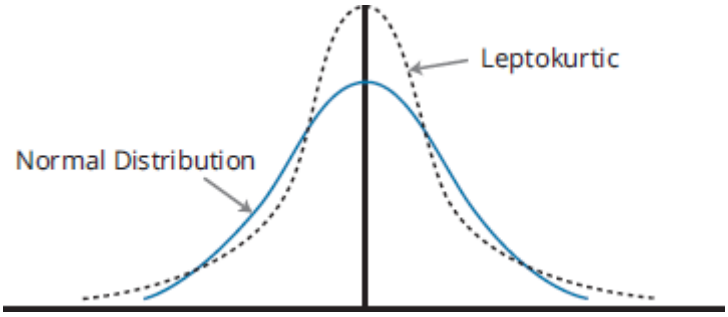
Note that the denominator is always positive, but that the numerator can be positive or negative depending on whether observations above the mean or observations below the mean tend to be farther from the mean, on average. When a distribution is right skewed, sample skewness is positive because the deviations above the mean are larger, on average. A left-skewed distribution has a negative sample skewness.

Dividing by standard deviation cubed standardizes the statistic and allows interpretation of the skewness measure. If relative skewness is equal to zero, the data are not skewed. Positive levels of relative skewness imply a positively skewed distribution, whereas negative values of relative skewness imply a negatively skewed distribution. Values of sample skewness in excess of 0.5 in absolute value are considered significant.

Kurtosis is a measure of the degree to which a distribution is more or less peaked than a normal distribution. **Leptokurtic** describes a distribution that is more peaked than a normal distribution, whereas **platykurtic** refers to a distribution that is less peaked, or flatter than a normal one. A distribution is **mesokurtic** if it has the same kurtosis as a normal distribution.

As indicated in Figure 3.3, a leptokurtic return distribution will have more returns clustered around the mean and more returns with large deviations from the mean (fatter tails). Relative to a normal distribution, a leptokurtic distribution will have a greater percentage of small deviations from the mean and a greater percentage of extremely large deviations from the mean. This means that there is a relatively greater probability of an observed value being either close to the mean or far from the mean. Regarding an investment returns distribution, a greater likelihood of a large deviation from the mean return is often perceived as an increase in risk.

Figure 3.3: Kurtosis



A distribution is said to exhibit **excess kurtosis** if it has either more or less kurtosis than the normal distribution. The computed kurtosis for all normal distributions is three. Statisticians, however, sometimes report excess kurtosis, which is defined as kurtosis minus three. Thus, a normal distribution has excess kurtosis equal to zero, a leptokurtic distribution has excess kurtosis greater than zero, and platykurtic distributions will have excess kurtosis less than zero.

Kurtosis is critical in a risk management setting. Most research about the distribution of securities returns has shown that returns are not normally distributed. Actual securities returns tend to exhibit both skewness and kurtosis. Skewness and kurtosis are critical concepts for risk management because when securities returns are modeled using an assumed normal distribution, the predictions from the models will not take into account the potential for extremely large, negative outcomes. In fact, most risk managers put very little emphasis on the mean and standard deviation of a distribution and focus more on the distribution of returns in the tails of the distribution—that is

where the risk is. In general, greater excess kurtosis and more negative skew in returns distributions indicate increased risk.

Sample kurtosis for large samples is approximated using deviations raised to the *fourth power*:

$$\text{sample kurtosis} \approx \left(\frac{1}{n}\right)\frac{\sum_{i=1}^{n}(X_i - \overline{X})^4}{s^4}$$

where:
s = sample standard deviation

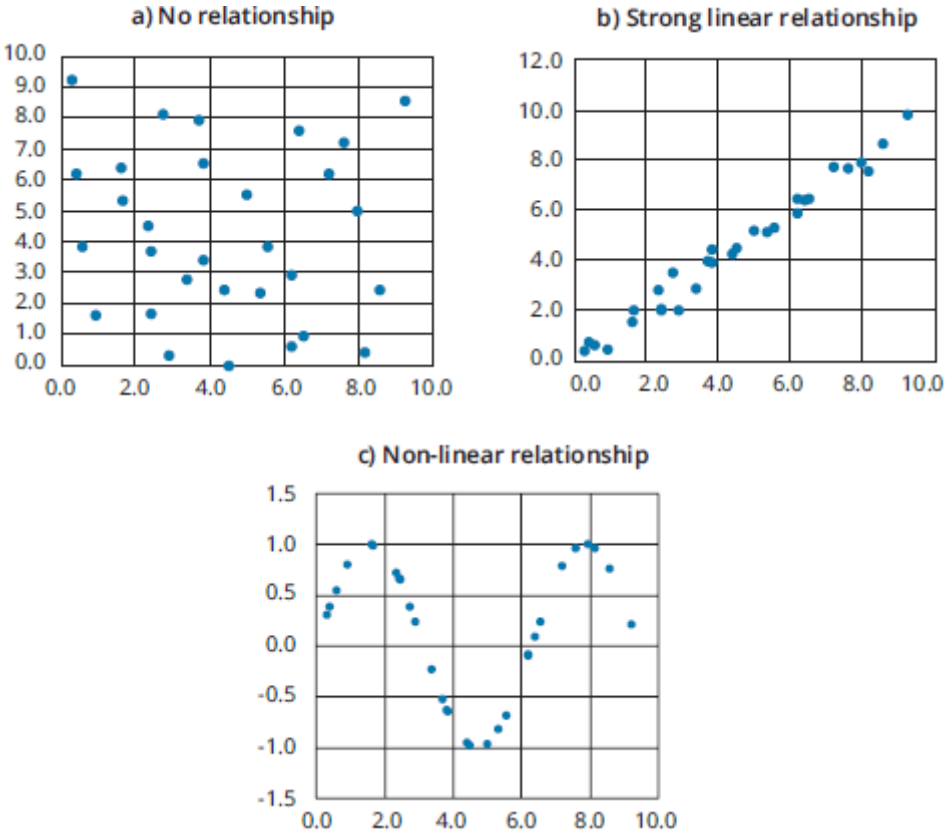
LOS 3.d: Interpret correlation between two variables to address an investment problem.

Scatter plots are a method for displaying the relationship between two variables. With one variable on the vertical axis and the other on the horizontal axis, their paired observations can each be plotted as a single point. For example, in Panel A of Figure 3.4, the point farthest to the upper right shows that when one of the variables (on the horizontal axis) equaled 9.2, the other variable (on the vertical axis) equaled 8.5.

The scatter plot in Panel A is typical of two variables that have no clear relationship. Panel B shows two variables that have a strong linear relationship—that is, a high correlation coefficient.

A key advantage of creating scatter plots is that they can reveal *nonlinear* relationships, which are not described by the correlation coefficient. Panel C illustrates such a relationship. Although the correlation coefficient for these two variables is close to zero, their scatter plot shows clearly that they are related in a predictable way.

Figure 3.4: Scatter Plots



Covariance is a measure of how two variables move together. The calculation of the **sample covariance** is based on the following formula:

$$s_{XY} = \frac{\sum_{i=1}^{n}\{[X_i - \overline{X}][Y_i - \overline{Y}]\}}{n-1}$$

where:
X_i = an observation of variable X
Y_i = an observation of variable Y
$\overline{X}$ = mean of variable X
$\overline{Y}$ = mean of variable Y
n = number of periods

In practice, the covariance is difficult to interpret. The value of covariance depends on the units of the variables. The covariance of daily price changes of two securities priced in yen will be much greater than their covariance if the securities are priced in dollars. Like the variance, the units of covariance are the square of the units used for the data.

Additionally, we cannot interpret the relative strength of the relationship between two variables. Knowing that the covariance of *X* and *Y* is 0.8756 tells us only that they tend to move together because the covariance is positive. A standardized measure of the linear relationship between two variables is called the **correlation coefficient**, or simply *correlation*. The correlation between two variables, *X* and *Y*, is calculated as follows:

$$\rho_{XY} = \frac{s_{XY}}{s_X s_Y}, \text{ which implies:}$$

$$s_{XY} = \rho_{XY} s_X s_Y$$

The *properties of the correlation* of two random variables, *X* and *Y*, are summarized here:

- Correlation measures the strength of the linear relationship between two random variables.
- Correlation has no units.
- The correlation ranges from −1 to +1. That is, $-1 \leq \rho_{XY} \leq +1$.
- If $\rho_{XY} = 1.0$, the random variables have perfect positive correlation. This means that a movement in one random variable results in a proportional positive movement in the other relative to its mean.
- If $\rho_{XY} = -1.0$, the random variables have perfect negative correlation. This means that a movement in one random variable results in an exact opposite proportional movement in the other relative to its mean.
- If $\rho_{XY} = 0$, there is no linear relationship between the variables, indicating that prediction of *Y* cannot be made on the basis of *X* using linear methods.

> **EXAMPLE: Correlation**
>
> The variance of returns on Stock A is 0.0028, the variance of returns on Stock B is 0.0124, and their covariance of returns is 0.0058. Calculate and interpret the correlation of the returns for Stocks A and B.
>
> **Answer:**
>
> First, it is necessary to convert the variances to standard deviations:
>
> $$s_A = (0.0028)^{1/2} = 0.0529$$
> $$s_B = (0.0124)^{1/2} = 0.1114$$
>
> Now, the correlation between the returns of Stock A and Stock B can be computed as follows:
>
> $$\rho_{AB} = \frac{0.0058}{(0.0529)(0.1114)} = 0.9842$$
>
> The fact that this value is close to +1 indicates that the linear relationship is not only positive, but also is very strong.

Care should be taken when drawing conclusions based on correlation. Causation is not implied just from significant correlation. Even if it were, which variable is causing change in the other is not revealed by correlation. It is more prudent to say that two variables exhibit positive (or negative) association, suggesting that the nature of any causal relationship is to be separately investigated or based on theory that can be subject to additional tests.

One question that can be investigated is the role of outliers (extreme values) in the correlation of two variables. If removing the outliers significantly reduces the

calculated correlation, further inquiry is necessary into whether the outliers provide information or are caused by noise (randomness) in the data used.

Spurious correlation refers to correlation that is either the result of chance or present due to changes in both variables over time that is caused by their association with a third variable. For example, we can find instances where two variables that are both related to the inflation rate exhibit significant correlation, but for which causation in either direction is not present.

In his book *Spurious Correlation*,[1] Tyler Vigen presents the following examples. The correlation between the age of each year's Miss America and the number of films Nicolas Cage appeared in that year is 87%. This seems a bit random. The correlation between the U.S. spending on science, space, and technology and suicides by hanging, strangulation, and suffocation over the 1999–2009 period is 99.87%. Impressive correlation, but both variables increased in an approximately linear fashion over the period.



MODULE QUIZ 3.2

1. Which of the following is *most accurate* regarding a distribution of returns that has a mean greater than its median?
   A. It is positively skewed.
   B. It is a symmetric distribution.
   C. It has positive excess kurtosis.
2. A distribution of returns that has a greater percentage of small deviations from the mean and a greater percentage of extremely large deviations from the mean compared with a normal distribution:
   A. is positively skewed.
   B. has positive excess kurtosis.
   C. has negative excess kurtosis.
3. The correlation between two variables is +0.25. The *most appropriate* way to interpret this value is to say:
   A. a scatter plot of the two variables is likely to show a strong linear relationship.
   B. when one variable is above its mean, the other variable tends to be above its mean as well.
   C. a change in one of the variables usually causes the other variable to change in the same direction.

KEY CONCEPTS

LOS 3.a

The arithmetic mean is the average of observations. The sample mean is the arithmetic mean of a sample:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

The median is the midpoint of a dataset when the data are arranged from largest to smallest.

The mode of a dataset is the value that occurs most frequently. The modal interval is a measure of mode for continuous data.

A trimmed mean omits outliers, and a winsorized mean replaces outliers with given values, reducing the effect of outliers on the mean in both cases.

Quantile is the general term for a value at or below which lies a stated proportion of the data in a distribution. Examples of quantiles include the following:

- *Quartile*. The distribution is divided into quarters.
- *Quintile*. The distribution is divided into fifths.
- *Decile*. The distribution is divided into tenths.
- *Percentile*. The distribution is divided into hundredths (percentages).

LOS 3.b

The range is the difference between the largest and smallest values in a dataset.

Mean absolute deviation (MAD) is the average of the absolute values of the deviations from the arithmetic mean:

$$\text{MAD} = \frac{\sum_{i=1}^{n} |X_i - \overline{X}|}{n}$$

Variance is defined as the mean of the squared deviations from the arithmetic mean:

$$\text{sample variance} = s^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n - 1}$$

where:

$\overline{X}$ = sample mean

n = sample size

Standard deviation is the positive square root of the variance, and it is frequently used as a quantitative measure of risk.

The coefficient of variation (CV) for sample data, $\text{CV} = \frac{s}{\overline{X}}$, is the ratio of the standard deviation of the sample to its mean.

Target downside deviation or semideviation is a measure of downside risk:

$$s_{\text{target}} = \sqrt{\frac{\sum_{\text{all } X_i < B}^{n} (X_i - B)^2}{n - 1}}$$

where B = the target

LOS 3.c

Skewness describes the degree to which a distribution is not symmetric about its mean. A right-skewed distribution has positive skewness. A left-skewed distribution has negative skewness.

For a positively skewed, unimodal distribution, the mean is greater than the median, which is greater than the mode. For a negatively skewed, unimodal distribution, the

mean is less than the median, which is less than the mode.

Kurtosis measures the peakedness of a distribution and the probability of extreme outcomes (thickness of tails):

- Excess kurtosis is measured relative to a normal distribution, which has a kurtosis of 3.
- Positive values of excess kurtosis indicate a distribution that is leptokurtic (fat tails, more peaked), so the probability of extreme outcomes is greater than for a normal distribution.
- Negative values of excess kurtosis indicate a platykurtic distribution (thin tails, less peaked).

LOS 3.d

Correlation is a standardized measure of association between two random variables. It ranges in value from −1 to +1 and is equal to $\frac{\text{Cov}_{A,B}}{\sigma_A \sigma_B}$.

Scatter plots are useful for revealing nonlinear relationships that are not measured by correlation.

Correlation does not imply that changes in one variable cause changes in the other. Spurious correlation may result by chance, or from the relationships of two variables to a third variable.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 3.1

1. **A** The winsorized mean substitutes a value for some of the largest and smallest observations. The trimmed mean removes some of the largest and smallest observations. (LOS 3.a)

2. **A** The sample mean is [22% + 5% + −7% + 11% + 2% + 11%] / 6 = 7.3%. The sample standard deviation is the square root of the sample variance:

$$s = \sqrt{\frac{(22 - 7.3)^2 + (5 - 7.3)^2 + (-7 - 7.3)^2 + (11 - 7.3)^2 + (2 - 7.3)^2 + (11 - 7.3)^2}{6 - 1}}$$
$$= \sqrt{96.3} = 9.8\%.$$

(LOS 3.b)

3. **A** Here are deviations from the target return:

$$22\% - 11\% = 11\%$$
$$5\% - 11\% = -6\%$$
$$-7\% - 11\% = -18\%$$
$$11\% - 11\% = 0\%$$
$$2\% - 11\% = -9\%$$
$$11\% - 11\% = 0\%$$

Target downside deviation =

$$\sqrt{\frac{(-6)^2 + (-18)^2 + (-9)^2}{6-1}} = \sqrt{88.2} = 9.39\%.$$

(LOS 3.b)

Module Quiz 3.2

1. **A** A distribution with a mean greater than its median is positively skewed, or skewed to the right. The skew pulls the mean. Kurtosis deals with the overall shape of a distribution, not its skewness. (LOS 3.c)

2. **B** A distribution that has a greater percentage of small deviations from the mean and a greater percentage of extremely large deviations from the mean will be leptokurtic and will exhibit excess kurtosis (positive). The distribution will be more peaked and have fatter tails than a normal distribution. (LOS 3.c)

3. **B** A correlation of +0.25 indicates a positive linear relationship between the variables—one tends to be above its mean when the other is above its mean. The value 0.25 indicates that the linear relationship is not particularly strong. Correlation does not imply causation. (LOS 3.d)

---

[1] "Spurious Correlations," Tyler Vigen, www.tylervigen.com

READING 4

PROBABILITY TREES AND CONDITIONAL EXPECTATIONS

MODULE 4.1: PROBABILITY MODELS, EXPECTED VALUES, AND BAYES' FORMULA

Video covering this content is available online.

LOS 4.a: Calculate expected values, variances, and standard deviations and demonstrate their application to investment problems.

The **expected value** of a random variable is the weighted average of the possible outcomes for the variable. The mathematical representation for the expected value of random variable *X*, that can take on any of the values from x_1 to x_n, is:

$$E(X) = \Sigma P(x_i)x_i = P(x_1)x_1 + P(x_2)x_2 + \ldots + P(x_n)x_n$$

where:

$P(x_i)$ = probability of outcome x_i

EXAMPLE: Expected earnings per share

The probability distribution of earnings per share (EPS) for Ron's Stores is given in the following figure. Calculate the expected EPS.

EPS Probability Distribution

Probability	EPS
10%	£1.80
20%	£1.60
40%	£1.20
30%	£1.00
100%	

Answer:

The expected EPS is simply a weighted average of each possible EPS, where the weights are the probabilities of each possible outcome:

$$E[EPS] = 0.10(1.80) + 0.20(1.60) + 0.40(1.20) + 0.30(1.00) = £1.28$$

Variance and *standard deviation* measure the dispersion of a random variable around its expected value, sometimes referred to as the **volatility** of a random variable. **Variance** (from a probability model) can be calculated as the probability-weighted sum of the squared deviations from the mean (or expected value). The standard deviation is the positive square root of the variance. The following example illustrates the calculations for a probability model of possible returns.

EXAMPLE: Expected value, variance, and standard deviation from a probability model

Using the probabilities given in the following table, calculate the expected return on Stock A, the variance of returns on Stock A, and the standard deviation of returns on Stock A.

Event	Prob	R_A	Prob × R_A	$R_A - E(R_A)$	$[R_A - E(R_A)]^2$	Prob × $[R_A - E(R_A)]^2$
Boom	30%	20%	0.06	0.07	0.0049	0.00147
Normal	50%	12%	0.06	−0.01	0.0001	0.00005
Slow	20%	5%	0.01	−0.08	0.0064	0.00128
			$E(R_A) = 0.13$			$Var(R_A) = 0.00280$

Answer:

$$E(R_A) = (0.30 \times 0.20) + (0.50 \times 0.12) + (0.20 \times 0.05) = 0.13 = 13\%$$

The expected return for Stock A is the probability-weighted sum of the returns under the three different economic scenarios.

In Column 5, we have calculated the differences between the returns under each economic scenario and the expected return of 13%.

In Column 6, we squared all the differences from Column 5. In the final column, we have multiplied the probabilities of each economic scenario times the squared deviation of returns from the expected returns, and their sum, 0.00280, is the variance of R_A.

The standard deviation of $R_A = \sqrt{0.0028} = 0.0529$.

Note that in a previous reading, we estimated the standard deviation of a distribution from sample data, rather than from a probability model of returns. For the sample standard deviation, we divided the sum of the squared deviations from the mean by n − 1, where *n* was the size of the sample. Here, we have no "n" because we have no observations; a probability model is forward-looking. We use the probability weights instead, as they describe the entire distribution of outcomes.

LOS 4.b: Formulate an investment problem as a probability tree and explain the use of conditional expectations in investment application.

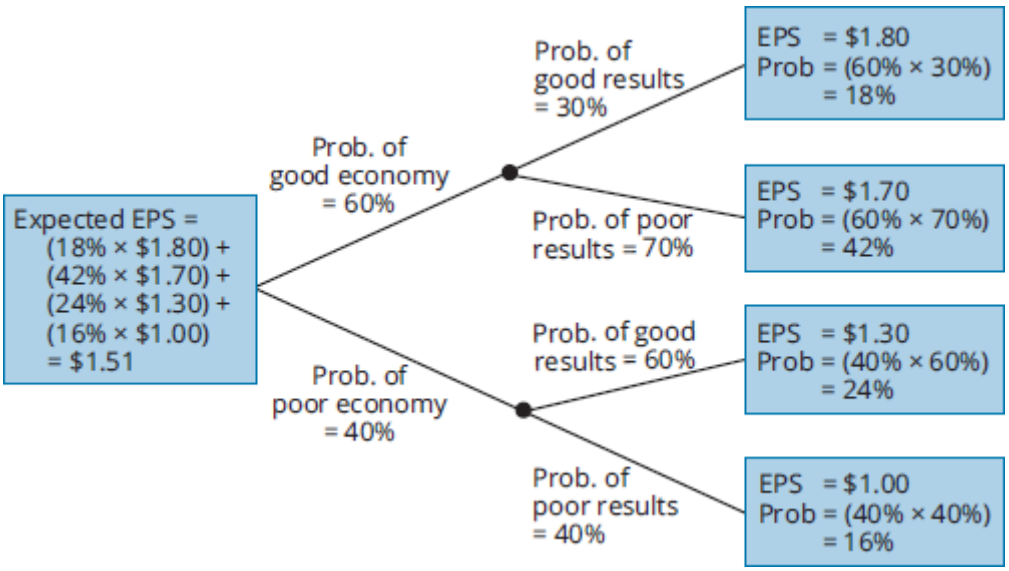
You may wonder where the returns and probabilities used in calculating expected values come from. A general framework, called a **probability tree**, is used to show the probabilities of various outcomes. In Figure 4.1, we have shown estimates of EPS for four different events: (1) a good economy and relatively good results at the company, (2) a good economy and relatively poor results at the company, (3) a poor economy and relatively good results at the company, and (4) a poor economy and relatively poor results at the company. Using the rules of probability, we can calculate the probabilities of each of the four EPS outcomes shown in the boxes on the right-hand side of the probability tree.

The expected EPS of $1.51 is simply calculated as follows:

$$(0.18 \times 1.80) + (0.42 \times 1.70) + (0.24 \times 1.30) + (0.16 \times 1.00) = \$1.51$$

Note that the probabilities of the four possible outcomes sum to 1.

Figure 4.1: A Probability Tree



Expected values or expected returns can be calculated using conditional probabilities. As the name implies, **conditional expected values** are contingent on the outcome of some other event. An analyst would use a conditional expected value to revise his expectations when new information arrives.

Consider the effect a tariff on steel imports might have on the returns of a domestic steel producer's stock in the previous example. The stock's conditional expected return, given that the government imposes the tariff, will be higher than the conditional expected return if the tariff is not imposed.

LOS 4.c: Calculate and interpret an updated probability in an investment setting using Bayes' formula.

Bayes' formula is used to update a given set of prior probabilities for a given event in response to the arrival of new information. The rule for updating prior probability of an event is as follows:

$$\text{updated probability} = \frac{\text{probability of new information for a given event}}{\text{unconditional probability of new information}} \times \text{prior probability of event}$$

We can derive Bayes' formula using the multiplication rule and noting that P(AB) = P(BA):

$$P(B \mid A) \times P(A) = P(BA), \text{ and } P(A \mid B) \times P(B) = P(AB)$$

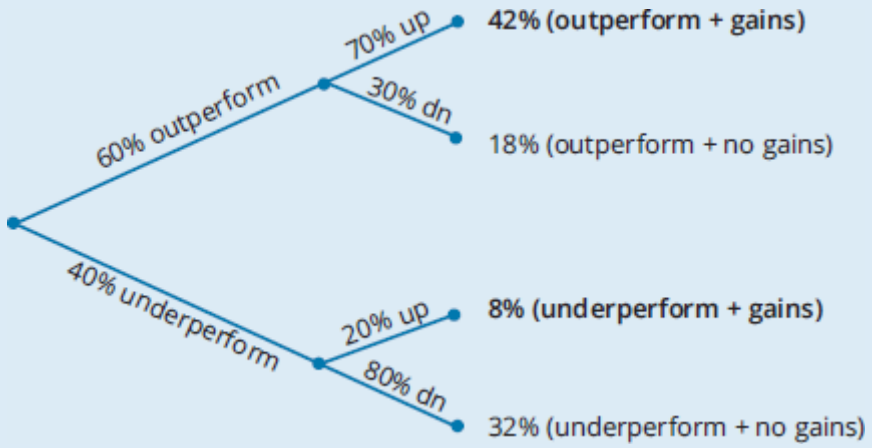
Because $P(BA) = P(AB)$, we can write $P(B \mid A)\,P(A) = P(A \mid B)\,P(B)$, and $\frac{P(B \mid A)P(A)}{P(B)}$ equals $\frac{P(BA)}{P(B)}$, the joint probability of A and B divided by the unconditional probability of B.

The following example illustrates the use of Bayes' formula. Note that A is *outperform* and A^C is *underperform*, P(BA) is (outperform + gains), $P(A^CB)$ is (underperform + gains), and the unconditional probability P(B) is P(AB) + $P(A^CB)$, by the total probability rule.

EXAMPLE: Bayes' formula

There is a 60% probability that the economy will outperform—and if it does, there is a 70% probability a stock will go up and a 30% probability the stock will go down. There is a 40% probability the economy will underperform, and if it does, there is a 20% probability the stock in question will increase in value (have gains) and an 80% probability it will not. Given that the stock increased in value, calculate the probability that the economy outperformed.

Answer:



In the figure just presented, we have multiplied the probabilities to calculate the probabilities of each of the four outcome pairs. Note that these sum to 1. Given that the stock has gains, what is our updated probability of an outperforming economy?

We sum the probability of stock gains in both states (outperform and underperform) to get 42% + 8% = 50%. Given that the stock has gains and using Bayes' formula, the probability that the economy has outperformed is $\frac{42\%}{50\%} = 84\%$.



MODULE QUIZ 4.1

1. Given the conditional probabilities in the following table and the unconditional probabilities P(Y = 1) = 0.3 and P(Y = 2) = 0.7, what is the expected value of *X*?

x_i	$P(x_i \mid Y = 1)$	$P(x_i \mid Y = 2)$
0	0.2	0.1
5	0.4	0.8
10	0.4	0.1

   A. 5.0.
   B. 5.3.
   C. 5.7.

2. An analyst believes that Davies Company has a 40% probability of earning more than $2 per share. She estimates that the probability that Davies Company's credit rating will be upgraded is 70% if its earnings per share (EPS) are greater than $2, and 20% if its EPS are $2 or less. Given the information that Davies Company's credit rating has been upgraded, what is the updated probability that its EPS are greater than $2?
   A. 50%.
   B. 60%.
   C. 70%.

KEY CONCEPTS

LOS 4.a

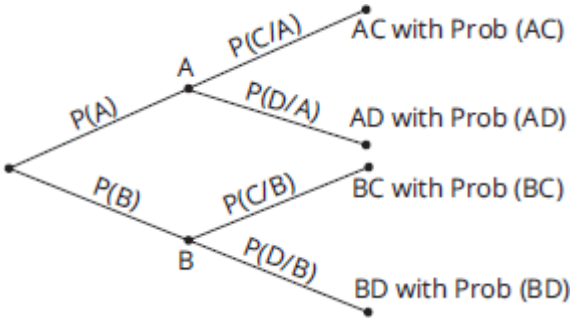
The expected value of a random variable is the weighted average of its possible outcomes:

$$E(X) = \Sigma P(x_i)x_i = P(x_1)x_1 + P(x_2)x_2 + \ldots + P(x_n)x_n$$

Variance can be calculated as the probability-weighted sum of the squared deviations from the mean or expected value. The standard deviation is the positive square root of the variance.

LOS 4.b

A probability tree shows the probabilities of two events and the conditional probabilities of two subsequent events:



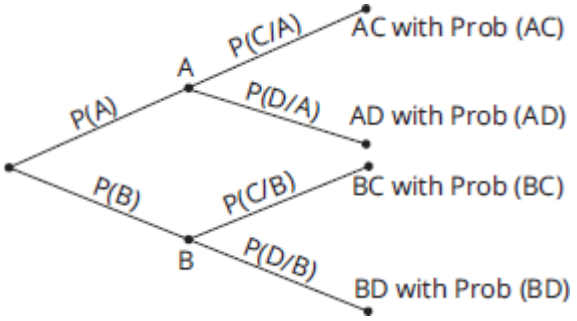
Conditional expected values depend on the outcome of some other event. Forecasts of expected values for a stock's return, earnings, and dividends can be refined, using conditional expected values, when new information arrives that affects the expected outcome.

LOS 4.c

Bayes' formula for updating probabilities based on the occurrence of an event *O* is as follows:

$$P(I|O) = \frac{P(O|I)}{P(O)} \times P(I)$$

Equivalently, based on the following tree diagram, $P(A|C) = \frac{P(AC)}{P(AC) + P(BC)}$:



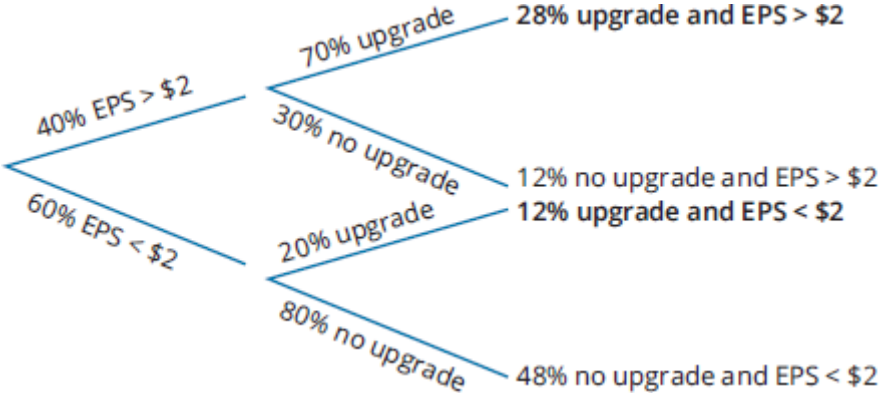
ANSWER KEY FOR MODULE QUIZZES

Module Quiz 4.1

1. **B** $E(X \mid Y = 1) = (0.2)(0) + (0.4)(5) + (0.4)(10) = 6$
   $E(X \mid Y = 2) = (0.1)(0) + (0.8)(5) + (0.1)(10) = 5$
   $E(X) = (0.3)(6) + (0.7)(5) = 5.30$

   (LOS 4.a)

2. **C** This is an application of Bayes' formula. As the following tree diagram shows, the updated probability that EPS are greater than \$2 is $\frac{28\%}{28\% + 12\%} = 70\%$:



   (LOS 4.c)

READING 5

PORTFOLIO MATHEMATICS

MODULE 5.1: PROBABILITY MODELS FOR PORTFOLIO RETURN AND RISK



Video covering this content is available online.

LOS 5.a: Calculate and interpret the expected value, variance, standard deviation, covariances, and correlations of portfolio returns.

The **expected return of a portfolio** composed of n assets with weights, w_i, and expected returns, R_i, can be determined using the following formula:

$$E(R_p) = \sum_{i=1}^{n} w_i E(R_i) = w_1 E(R_1) + w_2 E(R_2) + \ldots + w_n E(R_n)$$

The expected return and variance for a portfolio of assets can be determined using the properties of the individual assets in the portfolio. To do this, it is necessary to establish the portfolio weight for each asset. As indicated in the formula below, the weight, w, of Asset i is simply the market value currently invested in the asset divided by the current market value of the entire portfolio:

$$w_i = \frac{\text{market value of investment in Asset i}}{\text{market value of the portfolio}}$$

In many finance situations, we are interested in how two random variables move in relation to each other. For investment applications, one of the most frequently analyzed pairs of random variables is the returns of two assets. Investors and managers frequently ask questions such as, "What is the relationship between the return for Stock A and Stock B?" or, "What is the relationship between the performance of the S&P 500 and that of the automotive industry?"

Covariance is a measure of how two assets move together. It is the expected value of the product of the deviations of the two random variables from their respective expected values. A common symbol for the covariance between random variables X and Y is Cov(X,Y). Because we will be mostly concerned with the covariance of asset returns, the following formula has been written in terms of the covariance of the return of Asset i, R_i, and the return of Asset j, R_j:

$$\text{Cov}(R_i, R_j) = E\{[R_i - E(R_i)][R_j - E(R_j)]\}$$

The following are *properties of covariance*:

- The covariance of a random variable with itself is its variance; that is, $Cov(R_A,R_A) = Var(R_A)$.
- Covariance may range from negative infinity to positive infinity.
- A positive covariance indicates that when one random variable is above its mean, the other random variable also tends to be above its mean.
- A negative covariance indicates that when one random variable is above its mean, the other random variable tends to be below its mean.

The **sample covariance** for a sample of returns data can be calculated as follows:

$$s_{X,Y} = \frac{\sum_{i=1}^{n}\left\{\left[R_{1,i} - \overline{R_1}\right]\left[R_{2,i} - \overline{R_2}\right]\right\}}{n-1}$$

where:

$R_{1,i}$ = an observation of returns on Asset 1

$R_{2,i}$ = an observation of returns on Asset 2

$\overline{R_1}$ = mean return of Asset 1

$\overline{R_2}$ = mean return of Asset 2

n = number of observations in the sample

A **covariance matrix** shows the covariances between returns on a group of assets.

Figure 5.1: Covariance Matrix for Assets A, B, and C

Asset	A	B	C
A	$Cov(R_A,R_A)$	$Cov(R_A,R_B)$	$Cov(R_A,R_C)$
B	$Cov(R_B,R_A)$	$Cov(R_B,R_B)$	$Cov(R_B,R_C)$
C	$Cov(R_C,R_A)$	$Cov(R_C,R_B)$	$Cov(R_C,R_C)$

Note that the diagonal terms from top left are the variances of each asset's returns—in other words, $Cov(R_A,R_A) = Var(R_A)$.

The covariance between the returns on two assets does not depend on order—in other words, $Cov(R_A,R_B) = Cov(R_B,R_A)$—so in this covariance matrix, only three of the (off-diagonal) covariance terms are unique. In general for *n* assets, there are *n* variance terms (on the diagonal) and $n(n-1)/2$ unique covariance terms.

With **portfolio variance**, to calculate the variance of portfolio returns, we use the asset weights, returns variances, and returns covariances:

$$Var(R_P) = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j Cov(R_i,R_j)$$

The variance of a portfolio composed of two risky assets, A and B, can be expressed as follows:

$$Var(R_p) = w_A w_A Cov(R_A,R_A) + w_A w_B\, Cov(R_A,R_B) + w_B w_A\, Cov(R_B,R_A) + w_B w_B\, Cov(R_B,R_B)$$

We can write this more simply as:

$$\text{Var}(R_p) = w_A^2 \text{Var}(R_A) + w_B^2 \text{Var}(R_B) + 2w_A w_B \text{Cov}(R_A, R_B)$$

or:

$$\sigma_P^2 = w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + 2w_A w_B \text{Cov}_{AB}$$

For a three-asset portfolio, the portfolio variance is:

$$\sigma_P^2 = w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + w_C^2 \sigma_C^2 + 2w_A w_B \text{Cov}_{AB} + 2w_A w_C \text{Cov}_{AC} + 2w_B w_C \text{Cov}_{BC}$$

Consider a portfolio with three assets: an index of domestic stocks (60%), an index of domestic bonds (30%), and an index of international equities (10%). A covariance matrix of the three assets is shown here.

Figure 5.2: Covariance Matrix for the Three Assets

Asset	Domestic Stocks	Domestic Bonds	International Equities
Domestic stocks	400	44	180
Domestic bonds	44	70	35
International equities	180	35	450

Portfolio returns variance =

$$(0.6^2)400 + (0.3^2)70 + (0.1^2)450 + 2(0.6)(0.3)44 + 2(0.6)(0.1)180 + 2(0.3)(0.1)35 = 194.34$$

Portfolio returns standard deviation = $\sqrt{194.34} = 13.94\%$

Note that the units of variance and covariance are $\%^2$ (i.e., 0.001). When we put these values in as whole numbers (in $\%^2$), the portfolio variance is in $\%^2$, and the standard deviation is in whole percentages. We could also put variance and covariance in as decimals and get both the portfolio returns variance and standard deviation as decimals.

From the formula for portfolio returns variance, we can see that the lower the covariance terms, the lower the portfolio variance (and standard deviation). This is true for positive values of covariance, as well as negative values.

Recall that the correlation coefficient for two variables is:

$$\rho_{AB} = \frac{\text{Cov}_{AB}}{\sigma_A \sigma_B}\sigma, \text{ so that } (\text{Cov}_{AB}) = \rho_{AB} \times \sigma_A \sigma_B$$

This can be substituted for Cov_{AB} in our formula for portfolio returns variance. With this substitution, we can use a correlation matrix to calculate portfolio returns variance, rather than using covariances.

Figure 5.3: Correlation Matrix for the Three Assets

Asset	Domestic Stocks	Domestic Bonds	International Equities
Domestic stocks	1.000	0.263	0.424
Domestic bonds	0.263	1.000	0.197
International equities	0.424	0.197	1.000

Note that the correlations of asset returns with themselves (the diagonal terms) are all 1.

LOS 5.b: Calculate and interpret the covariance and correlation of portfolio returns using a joint probability function for returns.

EXAMPLE: Covariance of returns from a joint probability function

Assume that the economy can be in three possible states (S) next year: boom, normal, or slow economic growth. An expert source has calculated that P(boom) = 0.30, P(normal) = 0.50, and P(slow) = 0.20. The returns for Asset A, R_A, and Asset B, R_B, under each of the economic states are provided in the probability model as follows. What is the covariance of the returns for Asset A and Asset B?

Joint Probability Function

	$R_B = 30\%$	$R_B = 10\%$	$R_B = 0\%$
$R_A = 20\%$	0.30	0	0
$R_A = 12\%$	0	0.50	0
$R_A = 5\%$	0	0	0.20

The table gives us the joint probability of returns on Assets A and B (e.g., there is a 30% probability that the return on Asset A is 20% and the return on Asset B is 30%, and there is a 50% probability that the return on Asset A is 12% and the return on Asset B is 10%).

Answer:

First, we must calculate the expected returns for each of the assets:

$$E(R_A) = (0.3)(0.20) + (0.5)(0.12) + (0.2)(0.05) = 0.13$$

$$E(R_B) = (0.3)(0.30) + (0.5)(0.10) + (0.2)(0.00) = 0.14$$

The covariance can now be computed using the procedure described in the following table.

Covariance Calculation

Probability	R_A	R_B	Probability $\times [R_A - E(R_A)] \times [R_B - E(R_B)]$
0.3	0.20	0.30	$(0.3)(0.2 - 0.13)(0.3 - 0.14) = 0.00336$
0.5	0.12	0.10	$(0.5)(0.12 - 0.13)(0.1 - 0.14) = 0.00020$
0.2	0.05	0.00	$(0.2)(0.05 - 0.13)(0 - 0.14) = 0.00224$

The covariance of returns for Asset A and Asset B is 0.00336 + 0.00020 + 0.00224 = 0.0058.

LOS 5.c: Define shortfall risk, calculate the safety-first ratio, and identify an optimal portfolio using Roy's safety-first criterion.

Shortfall risk is the probability that a portfolio value or return will fall below a particular target value or return over a given period.

Roy's safety-first criterion states that the optimal portfolio minimizes the probability that the return of the portfolio falls below some minimum acceptable level. This minimum acceptable level is called the **threshold level**. Symbolically, Roy's safety-first criterion can be stated as follows:

minimize $P(R_p < R_L)$

where:

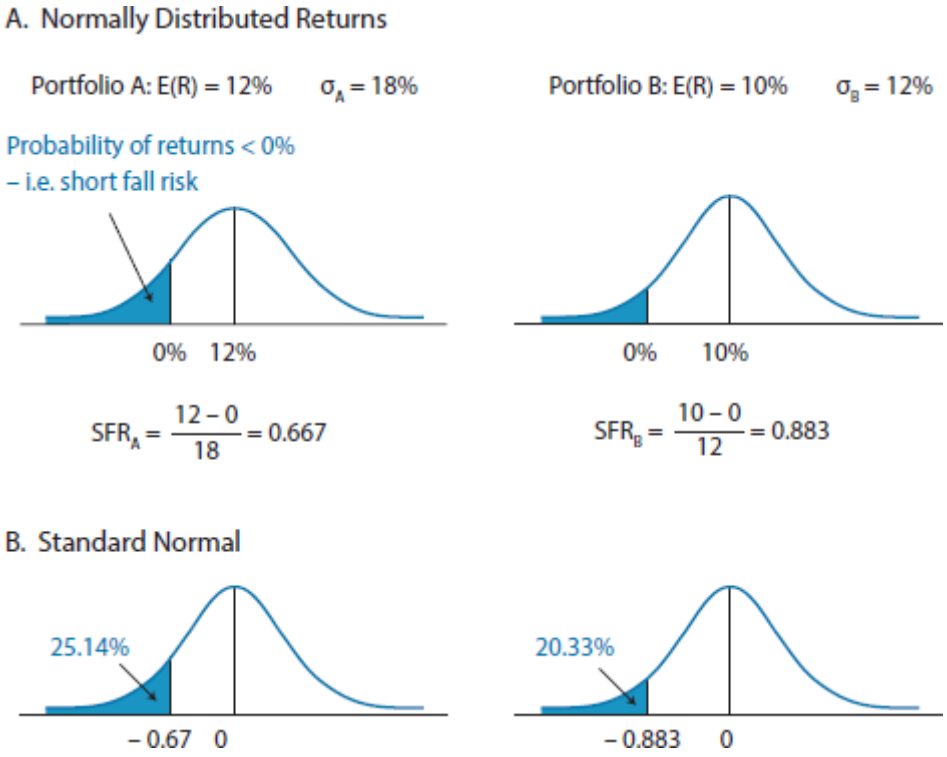
R_p = portfolio return
R_L = threshold level return

If portfolio returns are normally distributed, then Roy's safety-first criterion can be stated as follows:

maximize **safety-first ratio**, which equals $\frac{E(R_P) - R_L}{\sigma_P}$

The reasoning behind the safety-first criterion is illustrated in Figure 5.4. Assume an investor is choosing between two portfolios: Portfolio A, with an expected return of 12% and standard deviation of returns of 18%, and Portfolio B, with an expected return of 10% and standard deviation of returns of 12%. The investor has stated that he wants to minimize the probability of losing money (negative returns). Assuming that returns are normally distributed, the portfolio with the larger safety-first ratio using 0% as the threshold return (R_L) will be the one with the lower probability of negative returns.

Figure 5.4: The Safety-First Criterion and Shortfall Risk



Panel B of Figure 5.4 relates the safety-first ratio to the standard normal distribution. Note that the safety-first ratio is the number of standard deviations *below* the mean. Thus, the portfolio with the larger safety-first ratio has the lower probability of returns below the threshold return, which is a return of 0% in our example. Using a *z*-table for negative values, we can find the probabilities in the left-hand tails as indicated. These probabilities (25% for Portfolio A and 20% for Portfolio B) are also the shortfall risk for a target return of 0%—that is, the probability of negative returns. Portfolio B has the higher safety-first ratio, which means it has the lower probability of negative returns.

In summary, when choosing among portfolios with normally distributed returns using Roy's safety-first criterion, there are two steps:

Step 1: Calculate the safety-first ratio $= \frac{E(R_p) - R_L}{\sigma_p}$.

Step 2: Choose the portfolio that has the *largest* safety-first ratio.

EXAMPLE: Roy's safety-first criterion

For the next year, the managers of a $120 million college endowment plan have set a minimum acceptable end-of-year portfolio value of $123.6 million. Three portfolios are being considered that have the expected returns and standard deviation shown in the first two rows of the following table. Determine which of these portfolios is the most desirable using Roy's safety-first criterion and the probability that the portfolio value will fall short of the target amount.

Answer:

The threshold return is R_L = (123.6 − 120) / 120 = 0.030 = 3%. The safety-first ratios are shown in the following table. As indicated, the best choice is Portfolio A because it has the largest safety-first ratio.

Roy's Safety-First Ratios

Portfolio	Portfolio A	Portfolio B	Portfolio C
$E(R_p)$	9%	11%	6.6%
σ_p	12%	20%	8.2%
SFRatio	$0.5 = \frac{(9-3)}{12}$	$0.4 = \frac{(11-3)}{20}$	$0.44 = \frac{(6.6-3)}{8.2}$

The probability of an ending value for Portfolio A less than $123.6 million (a return less than 3%) is simply F(−0.5), which we can find on the *z*-table for negative values. The probability is 0.3085 = 30.85%.

MODULE QUIZ 5.1

1. The correlation of returns between Stocks A and B is 0.50. The covariance between these two securities is 0.0043, and the standard deviation of the return of Stock B is 26%. The variance of returns for Stock A is:
   A. 0.0011.
   B. 0.0331.
   C. 0.2656.

2. Given the following joint probability table for the returns on Assets P and Q:

	Q = 7%	Q = 4%	Q = 0%
P = 15%	0.2	0	0
P = 12%	0	0.2	0
P = 0%	0	0	0.6

   The covariance between P and Q is *closest* to:
   A. 18.0.
   B. 18.7.
   C. 19.3.

3. Expected returns and standard deviations of returns for three portfolios are as follows:

	Portfolio A	Portfolio B	Portfolio C
$E(R_p)$	5%	11%	18%
σ_p	8%	21%	40%

   Given a threshold level of return of 4%, the optimal portfolio using Roy's safety-first criterion is:
   A. Portfolio A.
   B. Portfolio B.

C. Portfolio C.

KEY CONCEPTS

LOS 5.a

The expected return of a portfolio composed of n assets with weights, w_i, and expected returns, R_i, is:

$$E(R_P) = \sum_{i=1}^{n} w_i E(R_i) = w_1 E(R_1) + w_2 E(R_2) + \ldots + w_n E(R_n)$$

The variance of a portfolio composed of two risky assets, A and B, can be expressed as follows:

$$\sigma_P^2 = w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + 2 w_A w_B Cov_{AB}$$

where Cov_{AB} is the expected value of the product of the deviations of the two assets' returns from their respective expected values.

The variance of a two-asset portfolio can also be expressed as follows:

$$Var(R_P) = w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + 2 w_A w_B \sigma_A \sigma_B \rho_{A,B}$$

where $\rho_{A,B}$ is the correlation of the two assets' returns.

LOS 5.b

Given the joint probabilities for A_i and B_i, the covariance is calculated as follows:

$$\sum_{i=1}^{n} P(A_i B_i)[A_i - E(A)][B_i - E(B)]$$

LOS 5.c

Shortfall risk is the probability that a portfolio's value (or return) will fall below a specific value over a given period.

The safety-first ratio for Portfolio P, based on a target return R_T, is:

$$SFRatio = \frac{E(R_P) - R_T}{\sigma_P}$$

Greater safety-first ratios are preferred and indicate a smaller shortfall probability. Roy's safety-first criterion states that the optimal portfolio minimizes shortfall risk.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 5.1

1. **A** $$\text{Corr}(R_A, R_B) = \frac{\text{Cov}(R_A, R_B)}{[\sigma(R_A)][\sigma(R_B)]}$$

$$\sigma^2(R_A) = \left[\frac{\text{Cov}(R_A, R_B)}{\sigma(R_B)\text{Corr}(R_A, R_B)}\right]^2$$

$$= \left[\frac{0.0043}{(0.26)(0.5)}\right]^2 = 0.0331^2 = 0.0011$$

(LOS 5.a)

2. **B** $\text{Expected value of P} = 0.2 \times 15 + 0.2 \times 12 + 0.6 \times 0 = 5.4\%$

$\text{Expected value of Q} = 0.2 \times 7 + 0.2 \times 4 + 0.6 \times 0 = 2.2\%$

$\text{Covariance} = 0.2 \times (15 - 5.4) \times (7 - 2.2) + 0.2 \times (12 - 5.4) \times (4 - 2.2) + 0.6\,(0 - 5.4)\,(0 - 2.2) = 18.72$

(LOS 5.b)

3. **C** Safety-first ratio for Portfolio A = (5 − 4) / 8 = 0.125

Safety-first ratio for Portfolio B = (11 − 4) / 21 = 0.300

Safety-first ratio for Portfolio C = (18 − 4) / 40 = 0.350

The largest value is 0.35, so Portfolio C has the smallest probability of a return below the threshold. (LOS 5.c)

READING 6

SIMULATION METHODS

MODULE 6.1: LOGNORMAL DISTRIBUTIONS AND SIMULATION TECHNIQUES

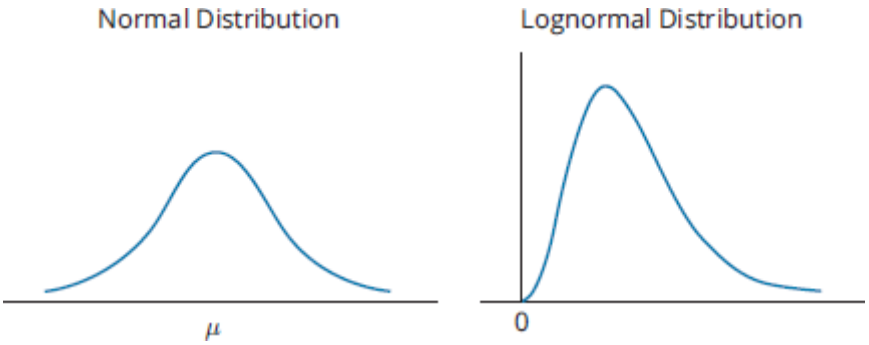
Video covering this content is available online.

LOS 6.a: Explain the relationship between normal and lognormal distributions and why the lognormal distribution is used to model asset prices when using continuously compounded asset returns.

The **lognormal distribution** is generated by the function e^x, where x is normally distributed. Because the natural logarithm, ln, of e^x is x, the logarithms of lognormally distributed random variables are normally distributed, thus the name.

Figure 6.1 illustrates the differences between a normal distribution and a lognormal distribution.

Figure 6.1: Normal vs. Lognormal Distributions



Recall from our reading about rates and returns that we use the natural logarithm to calculate continuously compounded returns. The lognormal distribution is useful for modeling asset prices if we think of an asset's future price as the result of a continuously compounded return on its current price. That is:

$$P_T = P_0 e^{r_{0,T}}$$

where:
P_T = asset price at time T
P_0 = asset price at time 0 (today)
$r_{0,T}$ = continuously compounded return on the asset from time 0 to time T

Because continuously compounded returns are additive, we can divide the time from 0 to T into shorter periods and state that $r_{0,T}$ is the sum of the continuously compounded returns over each of these shorter periods. Then, if we assume each of these returns is normally distributed, we can state that $r_{0,T}$ is normally distributed. Even if they are not normally distributed, the central limit theorem implies that their sum ($r_{0,T}$) is approximately normally distributed. This allows us to say P_T is lognormally distributed, because it is proportional to the logarithm of a normally distributed variable.

In many of the pricing models that we will see in the CFA curriculum, we assume returns are **independently and identically distributed**. If returns are *independently distributed*, past returns are not useful for predicting future returns. If returns are *identically distributed*, their mean and variance do not change over time (a property known as *stationarity* that is important in time series modeling, a Level II topic).

LOS 6.b: Describe Monte Carlo simulation and explain how it can be used in investment applications.

Monte Carlo simulation is a technique based on the repeated generation of one or more risk factors that affect security values to generate a distribution of security values. For each of the risk factors, the analyst must specify the parameters of the probability distribution that the risk factor is assumed to follow. A computer is then used to generate random values for each risk factor based on its assumed probability distributions. Each set of randomly generated risk factors is used with a pricing model to value the security. This procedure is repeated many times (100s, 1,000s, or 10,000s), and the distribution of simulated asset values is used to draw inferences about the expected (mean) value of the security—and possibly the variance of security values about the mean as well.

As an example, consider the valuation of stock options that can only be exercised on a particular date. The main risk factor is the value of the stock itself, but interest rates could affect the valuation as well. The simulation procedure would be to do the following:

1. Specify the probability distributions of stock prices and of the relevant interest rate, as well as the parameters (e.g., mean, variance, skewness) of the distributions.
2. Randomly generate values for both stock prices and interest rates.
3. Value the options for each pair of risk factor values.
4. After many iterations, calculate the mean option value and use that as your estimate of the option's value.

Monte Carlo simulation is used to do the following:

- Value complex securities.
- Simulate the profits/losses from a trading strategy.
- Calculate estimates of value at risk (VaR) to determine the riskiness of a portfolio of assets and liabilities.

- Simulate pension fund assets and liabilities over time to examine the variability of the difference between the two.
- Value portfolios of assets that have nonnormal return distributions.

An advantage of Monte Carlo simulation is that its inputs are not limited to the range of historical data. This allows an analyst to test scenarios that have not occurred in the past. The limitations of Monte Carlo simulation are that it is fairly complex and will provide answers that are no better than the assumptions about the distributions of the risk factors and the pricing/valuation model that is used. Also, simulation is not an analytic method, but a statistical one, and cannot offer the insights provided by an analytic method.

LOS 6.c: Describe the use of bootstrap resampling in conducting a simulation based on observed data in investment applications.

Resampling is another method for generating data inputs to use in a simulation. Often, we do not (or cannot) have data for a population, and can only approximate the population by sampling from it. (For example, we may think of the observed historical returns on an investment as a sample from the population of possible return outcomes.) To conduct resampling, we start with the observed sample and repeatedly draw subsamples from it, each with the same number of observations. From these samples, we can infer parameters for the population, such as its mean and variance.

In our reading on Estimation and Inference, we will describe some of the available resampling techniques. One of these is known as **bootstrap resampling**. In bootstrap resampling, we draw repeated samples of size *n* from the full dataset, replacing the sampled observations each time so that they might be redrawn in another sample. We can then directly calculate the standard deviation of these sample means as our estimate of the standard error of the sample mean.

Simulation using data from bootstrap resampling follows the same procedure as Monte Carlo simulation. The difference is the source and scope of the data. For example, if a simulation uses bootstrap resampling of historical returns data, its inputs are limited by the distribution of actual outcomes.



MODULE QUIZ 6.1

1. For a lognormal distribution the:
   A. mean equals the median.
   B. probability of a negative outcome is zero.
   C. probability of a positive outcome is 50%.
2. Which of the following is *least likely* to be a limitation of Monte Carlo analysis?
   A. Monte Carlo simulation is a statistical rather than an analytic method.
   B. Results of the analysis are no better than the assumptions used to generate it.
   C. Monte Carlo simulation is unable to provide answers to "what if" questions.
3. Which of the following is *most likely* a strength of bootstrapping?
   A. Offers a representation of the statistical features of a population.
   B. Provides only statistical estimates, not exact results.

C. Inputs may be limited by the distribution of actual outcomes.

KEY CONCEPTS

LOS 6.a

If x is normally distributed, e^x follows a lognormal distribution. The lognormal distribution is useful for modeling an asset's future price as the result of a continuously compounded return on its current price.

If investment returns are independently distributed, past returns are not useful for predicting future returns.

If investment returns are identically distributed, their mean and variance do not change over time.

LOS 6.b

Monte Carlo simulation uses randomly generated values for risk factors, based on their assumed distributions, to produce a distribution of possible security values. Its limitations are that it is fairly complex and will provide answers that are no better than the assumptions used.

LOS 6.c

Bootstrap resampling involves drawing repeated samples from a sample that represents the population, replacing the sampled observations each time so that they might be redrawn in another sample. The standard deviation of these sample means is an estimate of the standard error of the sample mean. As with all simulation techniques, its answers are no better than the assumptions used.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 6.1

1. **B** A lognormally distributed variable is never negative. (LOS 6.a)
2. **C** The ability to address "what if" questions (i.e., using input data outside the range that has been observed historically) is an advantage of Monte Carlo simulation. (LOS 6.b)
3. **A** One of the strengths of bootstrapping is that it offers a good representation of the statistical features of a population. However, this method does not provide exact results, and the inputs can be limited by the distribution of actual outcomes. (LOS 6.c)

READING 7

ESTIMATION AND INFERENCE

MODULE 7.1: SAMPLING TECHNIQUES AND THE CENTRAL LIMIT THEOREM



Video covering this content is available online.

LOS 7.a: Compare and contrast simple random, stratified random, cluster, convenience, and judgmental sampling and their implications for sampling error in an investment problem.

Probability sampling refers to selecting a sample when we know the probability of each sample member in the overall population. With **random sampling**, each item is assumed to have the same probability of being selected. If we have a population of data and select our sample by using a computer to randomly select a number of observations from the population, each data point has an equal probability of being selected—we call this **simple random sampling**. If we want to estimate the mean profitability for a population of firms, this may be an appropriate method.

Nonprobability sampling is based on either low cost and easy access to some data items, or on using the judgment of the researcher in selecting specific data items. Less randomness in selection may lead to greater sampling error.

Probability Sampling Methods

Simple random sampling is a method of selecting a sample in such a way that each item or person in the population being studied has the same likelihood of being included in the sample. As an example of simple random sampling, assume that you want to draw a sample of 5 items out of a group of 50 items. This can be accomplished by numbering each of the 50 items, placing them in a hat, and shaking the hat. Next, one number can be drawn randomly from the hat. Repeating this process (experiment) four more times results in a set of five numbers. The five drawn numbers (items) comprise a simple random sample from the population. In applications like this one, a random-number table or a computer random-number generator is often used to create the sample. Another way to form an approximately random sample is **systematic sampling**—selecting every *n*th member from a population.

Stratified random sampling uses a classification system to separate the population into smaller groups based on one or more distinguishing characteristics. From each subgroup, or stratum, a random sample is taken and the results are pooled. The size of

the samples from each stratum is based on the size of the stratum relative to the population.

Stratified sampling is often used in bond indexing because of the difficulty and cost of completely replicating the entire population of bonds. In this case, bonds in a population are categorized (stratified) according to major bond risk factors including, but not limited to, duration, maturity, and coupon rate. Then, samples are drawn from each separate category and combined to form a final sample.

To see how this works, suppose you want to construct a portfolio of 100 bonds that is indexed to a major municipal bond index of 1,000 bonds, using a stratified random sampling approach. First, the entire population of 1,000 municipal bonds in the index can be classified on the basis of maturity and coupon rate. Then, cells (stratum) can be created for different maturity/coupon combinations, and random samples can be drawn from each of the maturity/coupon cells. To sample from a cell containing 50 bonds with 2-to-4-year maturities and coupon rates less than 5%, we would select five bonds. The number of bonds drawn from a given cell corresponds to the cell's weight relative to the population (index), or (50/1,000) × 100 = 5 bonds. This process is repeated for all the maturity/coupon cells, and the individual samples are combined to form the portfolio.

By using stratified sampling, we guarantee that we sample five bonds from this cell. If we had used simple random sampling, there would be no guarantee that we would sample any of the bonds in the cell. Or, we may have selected more than five bonds from this cell.

Cluster sampling is also based on subsets of a population, but in this case, we are assuming that each subset (cluster) is representative of the overall population with respect to the item we are sampling. For example, we may have data on personal incomes for a state's residents by county. The data for each county is a cluster.

In **one-stage cluster sampling**, a random sample of clusters is selected, and all the data in those clusters comprise the sample. In **two-stage cluster sampling**, random samples from each of the selected clusters comprise the sample. Contrast this with stratified random sampling, in which random samples are selected from every subgroup.

To the extent that the subgroups do not have the same distribution as the entire population of the characteristic of interest, cluster sampling will have greater sampling error than simple random sampling. Two-stage cluster sampling can be expected to have greater sampling error than one-stage cluster sampling. Lower cost and less time required to assemble the sample are the primary advantages of cluster sampling, and it may be most appropriate for a smaller pilot study.

Nonprobability Sampling Methods

Convenience sampling refers to selecting sample data based on ease of access, using data that are readily available. Because such a sample is typically not random, sampling error will be greater. An analyst should initially look at the data before adopting a sampling method with less sampling error.

Judgmental sampling refers to samples for which each observation is selected from a larger dataset by the researcher, based on one's experience and judgment. As an example, a researcher interested in assessing company compliance with accounting standards may have experience suggesting that evidence of noncompliance is typically found in certain ratios derived from the financial statements. The researcher may select only data on these items. Researcher bias (or simply poor judgment) may lead to samples that have excessive sampling error. In the absence of bias or poor judgment, judgmental sampling may produce a more representative sample or allow the researcher to focus on a sample that offers good data on the characteristic or statistic of interest.

An important consideration when sampling is ensuring that the distribution of data of interest is constant for the whole population being sampled. For example, judging a characteristic of U.S. banks using data from 2005 to 2015 may not be appropriate. Regulatory reform of the banking industry after the financial crisis of 2007–2008 may have resulted in significant changes in banking practices, so that the mean of a statistic precrisis and its mean value across the population of banks postcrisis are quite different. Pooling the data over the entire period from 2005 to 2015 would not be appropriate if this is the case, and the sample mean calculated from these data would not be a good estimate of either precrisis or postcrisis mean values.

LOS 7.b: Explain the central limit theorem and its importance for the distribution and standard error of the sample mean.

The **central limit theorem** states that for simple random samples of size *n* from a population with a mean μ and a finite variance σ^2, the sampling distribution of the sample mean $\bar{x}$ approaches a normal probability distribution with mean μ and a variance equal to $\frac{\sigma^2}{n}$ as the sample size becomes large.

The central limit theorem is extremely useful because the normal distribution is relatively easy to apply to hypothesis testing and to the construction of confidence intervals. Specific inferences about the population mean can be made from the sample mean, *regardless of the population's distribution*, as long as the sample size is sufficiently large, which usually means $n \geq 30$.

Important properties of the central limit theorem include the following:

- If the sample size *n* is sufficiently large ($n \geq 30$), the sampling distribution of the sample means will be approximately normal. Remember what's going on here: random samples of size *n* are repeatedly being taken from an overall larger population. Each of these random samples has its own mean, which is itself a random variable, and this set of sample means has a distribution that is approximately normal.
- The mean of the population, μ, and the mean of the distribution of all possible sample means are equal.
- The variance of the distribution of sample means is $\frac{\sigma^2}{n}$, the population variance divided by the sample size.

The **standard error of the sample mean** is the standard deviation of the distribution of the sample means.

When the standard deviation of the population, σ, is *known*, the standard error of the sample mean is calculated as:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

where:
$\sigma_{\bar{x}}$ = standard error of the sample mean
σ = standard deviation of the population
n = size of the sample

However, practically speaking, the *population's standard deviation is almost never known*. Instead, the standard error of the sample mean must be estimated by dividing the standard deviation of *the sample* by $\sqrt{n}$:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

EXAMPLE: Standard error of sample mean (unknown population variance)

Suppose a sample contains the past 30 monthly returns for McCreary, Inc. The mean return is 2%, and the *sample* standard deviation is 20%. Calculate and interpret the standard error of the sample mean.

Answer:

Because σ is unknown, this is the standard error of the sample mean:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{20\%}{\sqrt{30}} = 3.6\%$$

This implies that if we took all possible samples of the size of 30 from McCreary's monthly returns and prepared a sampling distribution of the sample means, the mean would be 2% with a standard error of 3.6%.

EXAMPLE: Standard error of sample mean (unknown population variance)

Continuing with our example, suppose that instead of a sample size of 30, we take a sample of the past 200 monthly returns for McCreary, Inc. To highlight the effect of sample size on the sample standard error, let's assume that the mean return and standard deviation of this larger sample remain at 2% and 20%, respectively. Now, calculate the standard error of the sample mean for the 200-return sample.

Answer:

The standard error of the sample mean is computed as follows:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{20\%}{\sqrt{200}} = 1.4\%$$

The result of the preceding two examples illustrates an important property of sampling distributions. Notice that the value of the standard error of the sample mean decreased from 3.6% to 1.4% as the sample size increased from 30 to 200. This is because as the sample size increases, the sample mean gets closer, on average, to the true mean of the population. In other words, the distribution of the sample means about the population mean gets smaller and smaller, so the standard error of the sample mean decreases.

LOS 7.c: Describe the use of resampling (bootstrap, jackknife) to estimate the sampling distribution of a statistic.

Previously, we used the sample variance to calculate the standard error of our estimate of the mean. The standard error provides better estimates of the distribution of sample means when the sample is unbiased and the distribution of sample means is approximately normal.

Two alternative methods of estimating the standard error of the sample mean involve resampling of the data. The first of these, termed the **jackknife**, calculates multiple sample means, each with one of the observations removed from the sample. The standard deviation of these sample means can then be used as an estimate of the standard error of sample means. The jackknife is a computationally simple tool and can be used when the number of observations available is relatively small. This method can remove bias from statistical estimates.

The jackknife (so named because it is a handy and readily available tool) was developed when computational power was not as readily available and as low cost as today. A **bootstrap** method is more computationally demanding, but it has some advantages. To estimate the standard error of the sample mean, we draw repeated samples of size n from the full dataset (replacing the sampled observations each time). We can then directly calculate the standard deviation of these sample means as our estimate of the standard error of the sample mean.

The bootstrap method can improve accuracy compared to using only the data in a single sample, and it can be used to construct confidence intervals for various statistics in addition to the mean, such as the median. This method can also be used to estimate the distributions of complex statistics, including those that do not have an analytic form.



MODULE QUIZ 7.1

1. A simple random sample is a sample drawn in such a way that each member of the population has:
   A. some chance of being selected in the sample.
   B. an equal chance of being included in the sample.
   C. a 1% chance of being included in the sample.
2. To apply the central limit theorem to the sampling distribution of the sample mean, the sample is usually considered to be large if n is at least:
   A. 20.
   B. 25.

C. 30.

3. Which of the following techniques to improve the accuracy of confidence intervals on a statistic is *most* computationally demanding?
   A. Jackknife resampling.
   B. Systematic resampling.
   C. Bootstrap resampling.

KEY CONCEPTS

LOS 7.a

Simple random sampling is a method of selecting a sample in such a way that each item or person in the population being studied has the same probability of being included in the sample.

Stratified random sampling involves randomly selecting samples proportionally from subgroups that are formed based on one or more distinguishing characteristics of the data, so that random samples from the subgroups will have the same distribution of these characteristics as the overall population.

Cluster sampling is also based on subgroups (not necessarily based on data characteristics) of a larger dataset. In one-stage cluster sampling, the sample is formed from randomly chosen clusters (subsets) of the overall dataset. In two-stage cluster sampling, random samples are taken from each of the randomly chosen clusters (subgroups).

Convenience sampling refers to selecting sample data based on ease of access, using data that are readily available. Judgmental sampling refers to samples for which each observation is selected from a larger dataset by the researcher, based on the researcher's experience and judgment. Both are examples of nonprobability sampling and are nonrandom.

LOS 7.b

The central limit theorem states that for a population with a mean μ and a finite variance σ^2, the sampling distribution of the sample mean of all possible samples of size *n* (for $n \geq 30$) will be approximately normally distributed with a mean equal to μ and a variance equal to σ^2/n.

The standard error of the sample mean is the standard deviation of the distribution of the sample means and is calculated as $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ (where σ, the population standard deviation, is known) and as $s_{\overline{X}} = \frac{s}{\sqrt{n}}$ (where *s*, the sample standard deviation, is used because the population standard deviation is unknown).

LOS 7.c

Two resampling techniques to improve our estimates of the distribution of sample statistics are the jackknife and bootstrap. With the jackknife, we calculate *n* sample means, one with each observation in a sample of size *n* removed, and base our estimate on the standard error of sample means of size *n*. This can remove bias from our estimates based on the sample standard deviation without resampling.

With bootstrap resampling, we use the distribution of sample means (or other statistics) from a large number of samples of size *n*, drawn from a large dataset. Bootstrap resampling can improve our estimates of the distribution of various sample statistics and provide such estimates when analytical methods will not.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 7.1

1. **B** In a simple random sample, each element of the population has an equal probability of being selected. The 1% chance answer option allows for an equal chance, but only if there are 100 elements in the population from which the random sample is drawn. (LOS 7.a)

2. **C** Sample sizes of 30 or greater are typically considered large. (LOS 7.b)

3. **C** Bootstrap resampling, repeatedly drawing samples of equal size from a large dataset, is more computationally demanding than the jackknife. We have not defined *systematic resampling* as a specific technique. (LOS 7.c)

READING 8

HYPOTHESIS TESTING

MODULE 8.1: HYPOTHESIS TESTING BASICS

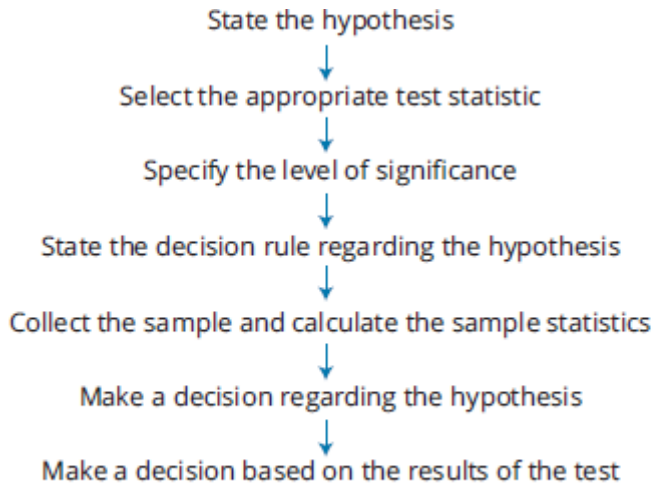
Video covering this content is available online.

LOS 8.a: Explain hypothesis testing and its components, including statistical significance, Type I and Type II errors, and the power of a test.

A hypothesis is a statement about the value of a population parameter developed for the purpose of testing a theory or belief. Hypotheses are stated in terms of the population parameter to be tested, like the population mean, μ. For example, a researcher may be interested in the mean daily return on stock options. Hence, the hypothesis may be that the mean daily return on a portfolio of stock options is positive.

Hypothesis testing procedures, based on sample statistics and probability theory, are used to determine whether a hypothesis is a reasonable statement and should not be rejected, or if it is an unreasonable statement and should be rejected. The process of hypothesis testing consists of a series of steps shown in Figure 8.1.

Figure 8.1: Hypothesis Testing Procedure



Source: Wayne W. Daniel and James C. Terrell, *Business Statistics, Basic Concepts and Methodology*, Houghton Mifflin, Boston, 1997.

The **null hypothesis**, designated H_0, is the hypothesis that the researcher wants to reject. It is the hypothesis that is actually tested and is the basis for the selection of the

test statistics. The null is generally stated as a simple statement about a population parameter. A typical statement of the null hypothesis for the population mean is H_0: $\mu = \mu_0$, where μ is the population mean and μ_0 is the hypothesized value of the population mean.

PROFESSOR'S NOTE
The null hypothesis always includes the "equal to" condition.

The **alternative hypothesis**, designated H_a, is what is concluded if there is sufficient evidence to reject the null hypothesis and is usually what you are really trying to assess. Why? You can never really prove anything with statistics—when the null hypothesis is discredited, the implication is that the alternative hypothesis is valid.

For a null hypothesis H_0: $\mu = \mu_0$, the alternative hypothesis is H_a: $\mu \neq \mu_0$. Notice that the null and alternative hypotheses are *mutually exclusive and exhaustive*. They include all possible outcomes of the test (they are exhaustive), and no possible outcome of the test satisfies both hypotheses (they are mutually exclusive).

Let's look at the development of the decision rule for a two-tailed test using a *z*-distributed test statistic (a *z*-test) at a 5% level of significance, $\alpha = 0.05$:

- At $\alpha = 0.05$, the computed test statistic is compared with the critical *z*-values of ±1.96. The values of ±1.96 correspond to $\pm z_{\alpha/2} = \pm z_{0.025}$, which is the range of *z*-values within which 95% of the probability lies. These values are obtained from the cumulative probability table for the standard normal distribution (*z*-table), which is included in the Appendix section of this book.
- If the computed test statistic falls outside the range of critical *z*-values (i.e., test statistic > 1.96, or test statistic < −1.96), we reject the null and conclude that the sample statistic is sufficiently different from the hypothesized value.
- If the computed test statistic falls within the range ±1.96, we conclude that the sample statistic is not sufficiently different from the hypothesized value ($\mu = \mu_0$, in this case), and we fail to reject the null hypothesis.

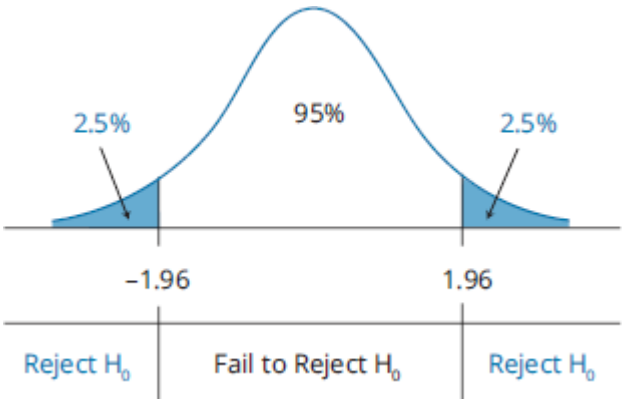
The **decision rule** (rejection rule) for a two-tailed *z*-test at $\alpha = 0.05$ can be stated as follows:

$$\text{Reject } H_0 \text{ if test statistic} < -1.96 \text{ or test statistic} > 1.96$$

Figure 8.2 shows the standard normal distribution for a two-tailed hypothesis test using the *z*-distribution. Notice that the significance level of 0.05 means that there is 0.05 / 2 = 0.025 probability (area) under each tail of the distribution beyond ±1.96.

Figure 8.2: Two-Tailed Hypothesis Test Using the Standard Normal (*z*)

Distribution



Hypothesis testing involves two statistics: (1) the test statistic calculated from the sample data and (2) the **critical value** of the test statistic. The value of the computed test statistic relative to the critical value is a key step in assessing the validity of a hypothesis.

PROFESSOR'S NOTE

Recall from the prerequisite readings that candidates are expected to know the most frequently used critical values for hypothesis tests with *z*-statistics:

Critical Value	Use for:
1.65	2-tailed test with 10% significance level, *or* 1-tailed test with 5% significance level
1.96	2 tailed test with 5% significance level
2.33	1-tailed test with 1% significance level
2.58	2-tailed test with 1% significance level

A test statistic is calculated by comparing the point estimate of the population parameter with the hypothesized value of the parameter (i.e., the value specified in the null hypothesis). As indicated in the following expression, the **test statistic** is the difference between the sample statistic and the hypothesized value, scaled by the standard error of the sample statistic:

$$\text{test statistic} = \frac{\text{sample statistic} - \text{hypothesized value}}{\text{standard error of the sample statistic}}$$

The standard error of the sample statistic is the adjusted standard deviation of the sample. When the sample statistic is the sample mean, $\bar{x}$, the standard error of the sample statistic for sample size n, is calculated as:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

when the population standard deviation, σ, *is known*, or:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

when the population standard deviation, σ, *is not known*. In this case, it is estimated using the standard deviation of the sample, *s*.

PROFESSOR'S NOTE
Don't be confused by the notation here. A lot of the literature you will encounter in your studies simply uses the term $\sigma_{\bar{x}}$ for the standard error of the test statistic, regardless of whether the population standard deviation or sample standard deviation was used in its computation.

As you will soon see, a test statistic is a random variable that may follow one of several distributions, depending on the characteristics of the sample and the population. We will look at four distributions for test statistics: the *t*-distribution, the *z*-distribution (standard normal distribution), the chi-square distribution, and the *F*-distribution. The critical value for the appropriate test statistic—the value against which the computed test statistic is compared—depends on its distribution.

Type I and Type II Errors

Keep in mind that hypothesis testing is used to make inferences about the parameters of a given population on the basis of statistics computed for a sample that is drawn from that population. We must be aware that there is some probability that the sample, in some way, does not represent the population, and any conclusion based on the sample about the population may be made in error.

When drawing inferences from a hypothesis test, there are two types of errors:

- **Type I error.** This is the rejection of the null hypothesis when it is actually true.
- **Type II error.** This is the failure to reject the null hypothesis when it is actually false.

The **significance level** is the probability of making a Type I error (rejecting the null when it is true) and is designated by the Greek letter alpha (α). For instance, a significance level of 5% (α = 0.05) means there is a 5% chance of rejecting a true null hypothesis. When conducting hypothesis tests, a significance level must be specified to identify the critical values needed to evaluate the test statistic.

While the significance level of a test is the probability of rejecting the null hypothesis when it is true, the **power of a test** is the probability of correctly rejecting the null hypothesis when it is false. The power of a test is actually one minus the probability of making a Type II error, or 1 − P(Type II error). In other words, the probability of rejecting the null when it is false (power of the test) equals one minus the probability of *not* rejecting the null when it is false (Type II error). When more than one test statistic may be used, the power of the test for the competing test statistics may be useful in deciding which test statistic to use. Ordinarily, we wish to use the test statistic that provides the most powerful test among all possible tests.

Figure 8.3 shows the relationship between the level of significance, the power of a test, and the two types of errors.

Figure 8.3: Type I and Type II Errors in Hypothesis Testing

Decision	True Condition	
	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Incorrect decision **Type II error**
Reject H_0	Incorrect decision **Type I error** Significance level, α, = P(Type I error)	Correct decision Power of the test = 1 − P(Type II error)

Sample size and the choice of significance level (Type I error probability) will together determine the probability of a Type II error. The relation is not simple, however, and calculating the probability of a Type II error in practice is quite difficult. Decreasing the significance level (probability of a Type I error) from 5% to 1%, for example, will increase the probability of failing to reject a false null (Type II error)—and, therefore, reduce the power of the test. Conversely, for a given sample size, we can increase the power of a test only with the cost that the probability of rejecting a true null (Type I error) increases. For a given significance level, we can decrease the probability of a Type II error and increase the power of a test only by increasing the sample size.

The decision for a hypothesis test is to either reject the null hypothesis or fail to reject the null hypothesis. Note that it is statistically incorrect to say "accept" the null hypothesis; it can only be supported or rejected. The **decision rule** for rejecting or failing to reject the null hypothesis is based on the distribution of the test statistic. For example, if the test statistic follows a normal distribution, the decision rule is based on critical values determined from the standard normal distribution (*z*-distribution). Regardless of the appropriate distribution, it must be determined if a one-tailed or two-tailed hypothesis test is appropriate before a decision rule (rejection rule) can be determined.

A decision rule is specific and quantitative. Once we have determined whether a one- or two-tailed test is appropriate, the significance level we require, and the distribution of the test statistic, we can calculate the exact critical value for the test statistic. Then, we have a decision rule of the following form: if the test statistic is (greater, less than) the value *X*, reject the null.

The ***p-value*** is the probability of obtaining a test statistic that would lead to a rejection of the null hypothesis, assuming the null hypothesis is true. It is the smallest level of significance for which the null hypothesis can be rejected.



MODULE QUIZ 8.1

1. For a hypothesis test with a probability of a Type II error of 60% and a probability of a Type I error of 5%, which of the following statements is *most accurate*?
   A. The power of the test is 40%, and there is a 5% probability that the test statistic will exceed the critical value(s).
   B. There is a 95% probability that the test statistic will be between the critical values, if this is a two-tailed test.

C. There is a 5% probability that the null hypothesis will be rejected when actually true, and the probability of rejecting the null when it is false is 40%.

2. If the significance level of a test is 0.05 and the probability of a Type II error is 0.15, what is the power of the test?
   A. 0.850.
   B. 0.950.
   C. 0.975.

MODULE 8.2: TYPES OF HYPOTHESIS TESTS



Video covering this content is available online.

LOS 8.b: Construct hypothesis tests and determine their statistical significance, the associated Type I and Type II errors, and power of the test given a significance level.

Here, we will show examples of hypothesis tests for population means and population variances:

- For a hypothesis concerning the *value of a population mean*, we use a *t*-test (or a *z*-test if the sample size is large enough).
- To test a hypothesis concerning the *equality of two population means*, we use a *t*-test. The nature of that test depends on whether the samples are independent (a difference in means test) or dependent (a paired comparisons test).
- For a hypothesis concerning the *value of a population variance*, we use a chi-square test.
- To test a hypothesis concerning the *equality of two population variances*, we use an *F*-test.

Value of a Population Mean

EXAMPLE: Hypothesis test of a population mean

A researcher has gathered data on the daily returns on a portfolio of call options over a recent 250-day period. The mean daily return has been 0.1%, and the sample standard deviation of daily portfolio returns is 0.25%. The researcher believes that the mean daily portfolio return is not equal to zero. Construct a hypothesis test of the researcher's belief.

Answer:

First, we need to specify the null and alternative hypotheses:

$$H_0: \mu_0 = 0 \text{ versus } H_a: \mu_0 \neq 0$$

The *t*-distribution is used to test a hypothesis about the value of a population mean. With 250 observations, however, this sample is considered to be large, so the *z*-distribution is acceptable. Because our sample is so large, the critical values for the *t*

and *z* are almost identical. Hence, there is almost no difference in the likelihood of rejecting a true null.

At a 5% level of significance, the critical *z*-values for a two-tailed test are ±1.96, so the decision rule can be stated as:

Reject H_0 if test statistic < −1.96 or test statistic > +1.96

Given a sample size of 250 with a standard deviation of 0.25%, the standard error can be computed as:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{0.25\%}{\sqrt{250}} = 0.0158\%$$

Using the standard error of the sample mean, our test statistic is:

$$\frac{0.001}{\left(\frac{0.0025}{\sqrt{250}}\right)} = \frac{0.001}{0.000158} = 6.33$$

Because 6.33 > 1.96, we can reject the null hypothesis that the mean daily option return is equal to zero.

Difference Between Means (Independent Samples)

We frequently want to know if there is a difference between the means of two populations. The *t*-test for **differences between means** requires that we are reasonably certain that our samples are independent and that they are taken from two populations that are normally distributed.



PROFESSOR'S NOTE

Please note the language of the LOS here. Candidates must "[c]onstruct hypothesis tests and determine their statistical significance …" Certainly, you should know that this is a *t*-test, and that we reject the hypothesis of equality when the test statistic is outside the critical *t*-values. Don't worry about memorizing the following formulas.

A pooled variance is used with the *t*-test for testing the hypothesis that the means of two normally distributed populations are equal, when the variances of the populations are unknown but assumed to be equal.

Assuming independent samples, the *t*-statistic is computed as follows:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

where:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

s_1^2 = variance of the first sample

s_2^2 = variance of the second sample

n_1 = number of observations in the first sample

n_2 = number of observations in the second sample

Note: degrees of freedom, *df*, is $(n_1 + n_2 - 2)$

Because we assume that the variances are equal, we just add the variances of the two sample means to calculate the standard error in the denominator.

The intuition here is straightforward. If the sample means are close together, the numerator of the *t*-statistic (and the *t*-statistic itself) are small, and we do not reject equality. If the sample means are far apart, the numerator of the *t*-statistic (and the *t*-statistic itself) are large, and we reject equality. Perhaps not as easy to remember is the fact that this test is only valid for two populations that are independent and normally distributed.

EXAMPLE: Difference between means—equal variances

Sue Smith is investigating whether the abnormal returns for acquiring firms during merger announcement periods differ for horizontal and vertical mergers. She estimates the abnormal returns for a sample of acquiring firms associated with horizontal mergers and a sample of acquiring firms involved in vertical mergers. Smith finds that abnormal returns from horizontal mergers have a mean of 1.0% and a standard deviation of 1.0%, while abnormal returns from vertical mergers have a mean of 2.5% and a standard deviation of 2.0%.

Smith assumes the samples are independent, the population means are normally distributed, and the population variances are equal.

Smith calculates the *t*-statistic as −5.474 and the degrees of freedom as 120. Using a 5% significance level, should Smith reject or fail to reject the null hypothesis that the abnormal returns to acquiring firms during the announcement period are the same for horizontal and vertical mergers?

Answer:

Because this is a two-tailed test, the structure of the hypotheses takes the following form:

$H_0: \mu_1 - \mu_2 = 0$ versus $H_a: \mu_1 - \mu_2 \neq 0$

where:

μ_1 = the mean of the abnormal returns for the horizontal mergers

μ_2 = the mean of the abnormal returns for the vertical mergers

From the following *t*-table segment, the critical *t*-value for a 5% level of significance at $\alpha / 2 = p = 0.025$ with df = 120, is 1.980.

Partial *t*-Table

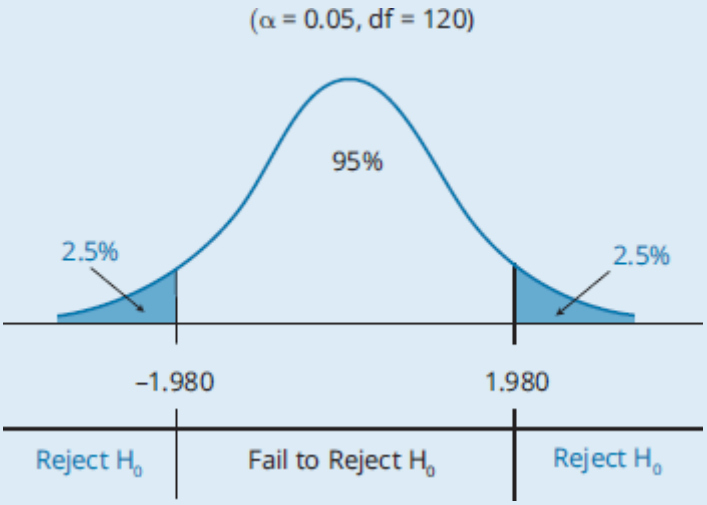
df	One-Tailed Probabilities (p)		
	p = 0.10	p = 0.05	p = 0.025
110	1.289	1.659	1.982
120	1.289	1.658	1.980
200	1.286	1.653	1.972

Thus, the decision rule can be stated as follows:

Reject H_0 if *t*-statistic < -1.980 or *t*-statistic > 1.980

The rejection region for this test is illustrated in the following figure.

Decision Rule for Two-Tailed *t*-Test



Because the test statistic, −5.474, falls to the left of the lower critical *t*-value, Smith can reject the null hypothesis and conclude that mean abnormal returns are different for horizontal and vertical mergers.

Paired Comparisons (Means of Dependent Samples)

While the test in the previous section was of the difference between the means of two independent samples, sometimes, our samples may be dependent. If the observations in the two samples both depend on some other factor, we can construct a **paired comparisons test** of whether the means of the differences between observations for the two samples are different. Dependence may result from an event that affects both

sets of observations for a number of companies, or because observations for two firms over time are both influenced by market returns or economic conditions.

For an example of a paired comparisons test, consider a test of whether the returns on two steel firms were equal over a five-year period. We can't use the difference in means test because we have reason to believe that the samples are not independent. To some extent, both will depend on the returns on the overall market (market risk) and the conditions in the steel industry (industry-specific risk). In this case, our pairs will be the returns on each firm over the same time periods, so we use the differences in monthly returns for the two companies. The paired comparisons test is just a test of whether the average difference between monthly returns is significantly different from some value, typically zero, based on the standard error of the differences in monthly returns.

Remember, the paired comparisons test also requires that the sample data be normally distributed. Although we frequently just want to test the hypothesis that the mean of the differences in the pairs is zero ($\mu_{dz} = 0$), the general form of the test for any hypothesized mean difference, μ_{dz}, is as follows:

$$H_0: \mu_d = \mu_{dz} \text{ versus } H_a: \mu_d \neq \mu_{dz}$$

where:

μ_d = mean of the population of paired differences

μ_{dz} = hypothesized mean of paired differences, which is commonly zero

For one-tail tests, the hypotheses are structured as either of these:

$$H_0: \mu_d \leq \mu_{dz} \text{ versus } H_a: \mu_d > \mu_{dz}, \text{ or } H_0: \mu_d \geq \mu_{dz} \text{ versus } H_a: \mu_d < \mu_{dz}$$

For the paired comparisons test, the *t*-statistic with n – 1 degrees of freedom is computed as follows:

$$t = \frac{\bar{d} - \mu_{dz}}{s_{\bar{d}}}$$

where:

$\bar{d}$ = sample mean difference = $\frac{1}{n}\sum_{i=1}^{n} d_i$

d_i = difference between the *i*th pair of observations

$s_{\bar{d}}$ = standard error of the mean difference = $\frac{s_d}{\sqrt{n}}$

s_d = sample standard deviation = $\sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}}$

n = number of paired observations

EXAMPLE: Paired comparisons test

Joe Andrews is examining changes in estimated betas for the common stock of companies in the telecommunications industry before and after deregulation. Andrews believes the betas may decline because of deregulation, because companies are no longer subject to the uncertainties of rate regulation—or that they may

increase because there is more uncertainty regarding competition in the industry. Andrews calculates a *t*-statistic of 10.26 for this hypothesis test, based on a sample size of 39. Using a 5% significance level, determine whether there is a change in betas.

Answer:

Because the mean difference may be positive or negative, a two-tailed test is in order here. Thus, the hypotheses are structured as:

$$H_0: \mu_d = 0 \text{ versus } H_a: \mu_d \neq 0$$

There are 39 − 1 = 38 degrees of freedom. Using the *t*-distribution, the two-tailed critical *t*-values for a 5% level of significance with df = 38 is ±2.024. As indicated in the following table, the critical *t*-value of 2.024 is located at the intersection of the p = 0.025 column and the df = 38 row. The one-tailed probability of 0.025 is used because we need 2.5% in each tail for 5% significance with a two-tailed test.

Partial *t*-Table

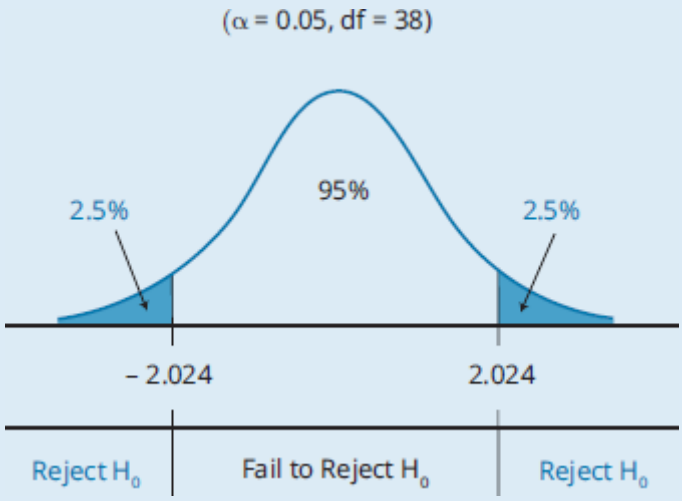
df	One-Tailed Probabilities (p)		
	p = 0.10	p = 0.05	p = 0.025
38	1.304	1.686	2.024
39	1.304	1.685	2.023
40	1.303	1.684	2.021

Thus, this is the decision rule:

$$\text{Reject } H_0 \text{ if } t\text{-statistic} < -2.024, \text{ or } t\text{-statistic} > 2.024$$

This decision rule is illustrated in the following figure.

Decision Rule for a Two-Tailed Paired Comparisons Test



The test statistic, 10.26, is greater than the critical *t*-value, 2.024—it falls in the rejection region to the right of 2.024 in the previous figure. Thus, we reject the null

hypothesis of no difference, concluding that there *is* a statistically significant difference between mean firm betas before and after deregulation.

Keep in mind that we have been describing two distinct hypothesis tests: one about the significance of the difference between the means of two populations and one about the significance of the mean of the differences between pairs of observations. Here are rules for when these tests may be applied:

- The test of the differences in means is used when there are two *independent samples*.
- A test of the significance of the mean of the differences between paired observations is used when the samples are *not independent*.



PROFESSOR'S NOTE
Again, the LOS here says "[c]onstruct hypothesis tests and determine their statistical significance …" We can't believe candidates are expected to memorize these formulas (or that you would be a better analyst if you did). You should instead focus on the fact that both of these tests involve *t*-statistics and depend on the degrees of freedom. Also note that when samples are independent, you can use the difference in means test, and when they are dependent, we must use the paired comparison (mean differences) test. In that case, with a null hypothesis that there is no difference in means, the test statistic is simply the mean of the differences between each pair of observations, divided by the standard error of those differences. This is just a straightforward *t*-test of whether the mean of a sample is zero, which might be considered fair game for the exam.

Value of a Population Variance

The *chi-square test* is used for hypothesis tests concerning the variance of a normally distributed population. Letting σ^2 represent the true population variance and σ_0^2 represent the hypothesized variance, the hypotheses for a two-tailed test of a single population variance are structured as follows:

$$H_0: \sigma^2 = \sigma_0^2 \text{ versus } H_a: \sigma^2 \neq \sigma_0^2$$

The hypotheses for one-tailed tests are structured as follows:

$$H_0: \sigma^2 \leq \sigma_0^2 \text{ versus } H_a: \sigma^2 > \sigma_0^2 \text{ or } H_0: \sigma^2 \geq \sigma_0^2 \text{ versus } H_a: \sigma^2 < \sigma_0^2$$

Hypothesis testing of the population variance requires the use of a chi-square distributed test statistic, denoted χ^2. The chi-square distribution is asymmetrical and approaches the normal distribution in shape as the degrees of freedom increase.

To illustrate the chi-square distribution, consider a two-tailed test with a 5% level of significance and 30 degrees of freedom. As displayed in Figure 8.4, the critical chi-square values are 16.791 and 46.979 for the lower and upper bounds, respectively. These values are obtained from a chi-square table, which is used in the same manner as a *t*-table. A portion of a chi-square table is presented in Figure 8.5.

Note that the chi-square values in Figure 8.5 correspond to the probabilities in the right tail of the distribution. As such, the 16.791 in Figure 8.4 is from the column headed *0.975* because 95% + 2.5% of the probability is to the right of it. The 46.979 is from the column headed *0.025* because only 2.5% probability is to the right of it. Similarly, at a 5% level of significance with 10 degrees of freedom, Figure 8.5 shows that the critical chi-square values for a two-tailed test are 3.247 and 20.483.

Figure 8.4: Decision Rule for a Two-Tailed Chi-Square Test

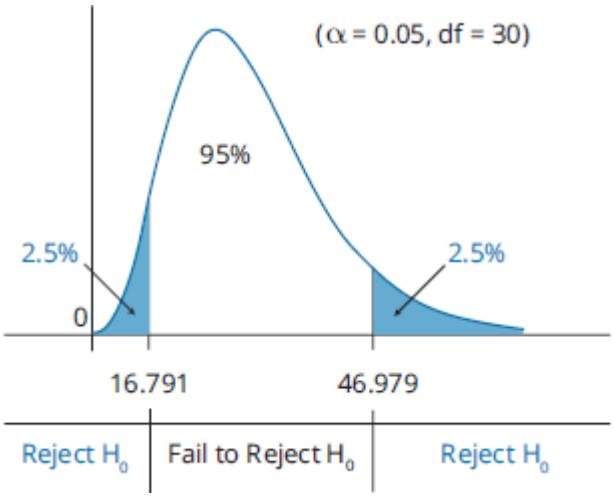


Figure 8.5: Chi-Square Table

	Probability in Right Tail					
Degrees of Freedom	0.975	0.95	0.90	0.1	0.05	0.025
9	2.700	3.325	4.168	14.684	16.919	19.023
10	3.247	3.940	4.865	15.987	18.307	20.483
11	3.816	4.575	5.578	17.275	19.675	21.920
30	16.791	18.493	20.599	40.256	43.773	46.979

The chi-square test statistic, χ^2, with n − 1 degrees of freedom, is computed as follows:

$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma_0^2}$$

where:

n = sample size

s^2 = sample variance

σ_0^2 = hypothesized value for the population variance

Similar to other hypothesis tests, the chi-square test compares the test statistic, χ^2_{n-1}, to a critical chi-square value at a given level of significance and n − 1 degrees of freedom. Because the chi-square distribution is bounded below by zero, chi-square values cannot be negative.

EXAMPLE: Chi-square test for a single population variance

Historically, the High-Return Equity Fund has advertised that its monthly returns have a standard deviation equal to 4%. This was based on estimates from the 2005–2013 period. High-Return wants to verify whether this claim still adequately describes the standard deviation of the fund's returns. High-Return collected monthly returns for the 24-month period between 2013 and 2015 and measured a standard deviation of monthly returns of 3.8%. High-Return calculates a test statistic of 20.76. Using a 5% significance level, determine if the more recent standard deviation is different from the advertised standard deviation.

Answer:

The null hypothesis is that the standard deviation is equal to 4% and, therefore, the variance of monthly returns for the population is $(0.04)^2 = 0.0016$. Because High-Return simply wants to test whether the standard deviation has changed, up or down, a two-sided test should be used. The hypothesis test structure takes this form:

$$H_0: \sigma_0^2 = 0.0016 \text{ versus } H_a: \sigma^2 \neq 0.0016$$

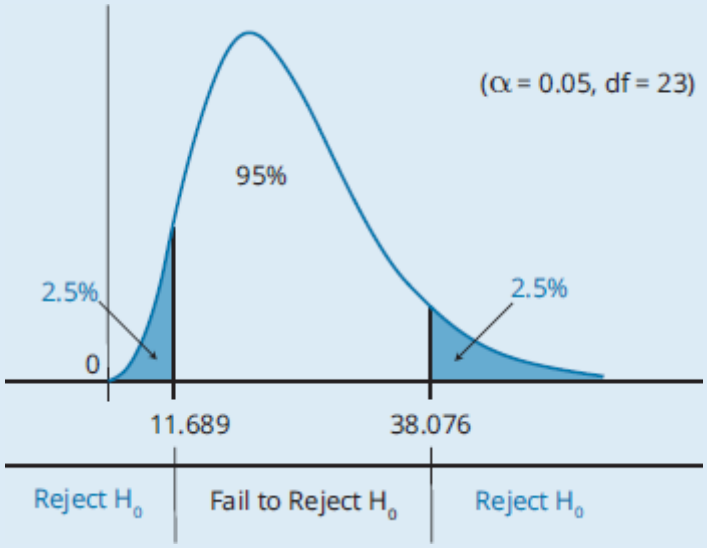
The appropriate test statistic for tests of variance is a chi-square statistic.

With a 24-month sample, there are 23 degrees of freedom. Using the table of chi-square values in Appendix E of this book, for 23 degrees of freedom and probabilities of 0.975 and 0.025, we find two critical values, 11.689 and 38.076. Thus, this is the decision rule:

$$\text{Reject } H_0 \text{ if } \chi^2 < 11.689, \text{ or } \chi^2 > 38.076$$

This decision rule is illustrated in the following figure.

Decision Rule for a Two-Tailed Chi-Square Test of a Single Population Variance



Because the computed test statistic, χ^2, falls between the two critical values, we cannot reject the null hypothesis that the variance is equal to 0.0016. The recently measured standard deviation is close enough to the advertised standard deviation that we cannot say that it is different from 4%, at a 5% level of significance.

Comparing Two Population Variances

The hypotheses concerned with the equality of the variances of two populations are tested with an *F*-distributed test statistic. Hypothesis testing using a test statistic that follows an *F*-distribution is referred to as the *F*-test. The *F*-test is used under the assumption that the populations from which samples are drawn are normally distributed, and that the samples are independent.

If we let σ_1^2 and σ_2^2 represent the variances of normal Population 1 and Population 2, respectively, the hypotheses for the two-tailed *F*-test of differences in the variances can be structured as follows:

$$H_0: \sigma_1^2 = \sigma_2^2 \text{ versus } H_a: \sigma_1^2 \neq \sigma_2^2$$

The one-sided test structures can be specified as follows:

$$H_0: \sigma_1^2 \leq \sigma_2^2 \text{ versus } H_a: \sigma_1^2 > \sigma_2^2, \text{ or } H_0: \sigma_1^2 \geq \sigma_2^2 \text{ versus } H_a: \sigma_1^2 < \sigma_2^2$$

The test statistic for the *F*-test is the ratio of the sample variances. The *F*-statistic is computed as follows:

$$F = \frac{s_1^2}{s_2^2}$$

where:

S_1^2 = variance of the sample of n_1 observations drawn from Population 1

S_2^2 = variance of the sample of n_2 observations drawn from Population 2

Note that $n_1 - 1$ and $n_2 - 1$ are the degrees of freedom used to identify the appropriate critical value from the *F*-table (provided in this book's Appendix).

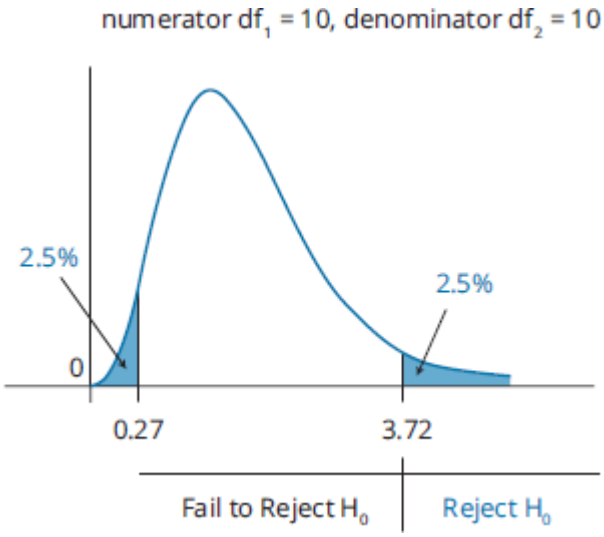


PROFESSOR'S NOTE

Always put the larger variance in the numerator (s_1^2). Following this convention means that we only have to consider the critical value for the right-hand tail.

An ***F-distribution*** is presented in Figure 8.6. As indicated, the *F*-distribution is right skewed and is bounded by zero on the left-hand side. The shape of the *F*-distribution is determined by *two separate degrees of freedom*: the numerator degrees of freedom, df_1, and the denominator degrees of freedom, df_2.

Figure 8.6: *F*-Distribution



Note that when the sample variances are equal, the value of the test statistic is 1. The upper critical value is always greater than one (the numerator is significantly greater than the denominator), and the lower critical value is always less than one (the numerator is significantly smaller than the denominator). In fact, the lower critical value is the reciprocal of the upper critical value. For this reason, in practice, we put the larger sample variance in the numerator and consider only the upper critical value.

EXAMPLE: *F*-test for equal variances

Annie Cower is examining the earnings for two different industries. Cower suspects that the variance of earnings in the textile industry is different from the variance of earnings in the paper industry. To confirm this suspicion, Cower has looked at a sample of 31 textile manufacturers and a sample of 41 paper companies. She measured the sample standard deviation of earnings across the textile industry to be $4.30, and that of the paper industry companies to be $3.80. Cower calculates a test statistic of 1.2805. Using a 5% significance level, determine if the earnings of the textile industry have a different standard deviation than those of the paper industry.

Answer:

In this example, we are concerned with whether the variance of earnings for companies in the textile industry is equal to the variance of earnings for companies in the paper industry. As such, the test hypotheses can be appropriately structured as follows:

$$H_0: \sigma_1^2 = \sigma_2^2 \text{ versus } H_a: \sigma_1^2 \neq \sigma_2^2$$

For tests of difference between variances, the appropriate test statistic is:

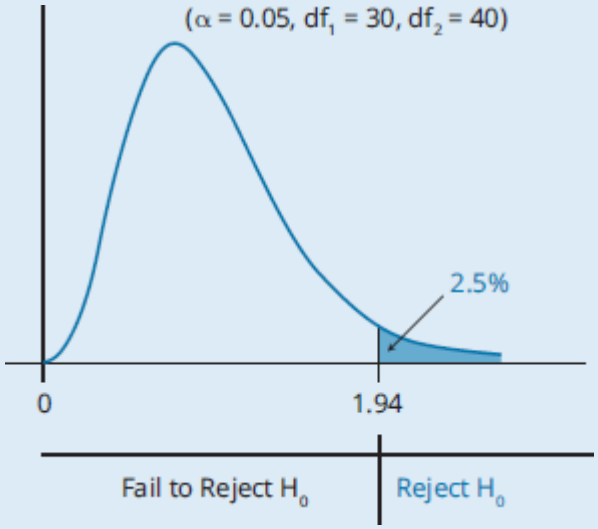
$$F = \frac{s_1^2}{s_2^2}$$

where s_1^2 is the larger sample variance.

Using the sample sizes for the two industries, the critical *F*-value for our test is found to be 1.94. This value is obtained from the table of the *F*-distribution for 2.5% in the upper tail, with $df_1 = 30$ and $df_2 = 40$. Thus, if the computed *F*-statistic is greater than the critical value of 1.94, the null hypothesis is rejected. The decision rule, illustrated in the following figure, can be stated as follows:

Reject H_0 if $F > 1.94$

Decision Rule for *F*-Test



Because the calculated *F*-statistic of 1.2805 is less than the critical *F*-statistic of 1.94, Cower cannot reject the null hypothesis. Cower should conclude that the earnings variances of the industries are not significantly different from one another at a 5% level of significance.

LOS 8.c: Compare and contrast parametric and nonparametric tests, and describe situations where each is the more appropriate type of test.

Parametric tests rely on assumptions regarding the distribution of the population and are specific to population parameters. For example, the *z*-test relies upon a mean and a standard deviation to define the normal distribution. The *z*-test also requires that either the sample is large, relying on the central limit theorem to assure a normal sampling distribution, or that the population is normally distributed.

Nonparametric tests either do not consider a particular population parameter or have few assumptions about the population that is sampled. Nonparametric tests are used when there is concern about quantities other than the parameters of a distribution, or when the assumptions of parametric tests can't be supported. They are also used when the data are not suitable for parametric tests (e.g., ranked observations).

Some situations where a nonparametric test is called for may include the following:

1. The assumptions about the distribution of the random variable that support a parametric test are not met. An example would be a hypothesis test of the mean

value for a variable that comes from a distribution that is not normal and is of small size, so that neither the *t*-test nor the *z*-test is appropriate.

2. A nonparametric test is called for when data are ranks (an ordinal measurement scale) rather than values.
3. The hypothesis does not involve the parameters of the distribution, such as testing whether a variable is normally distributed. We can use a nonparametric test, called a runs test, to determine whether data are random. A runs test provides an estimate of the probability that a series of changes (e.g., +, +, −, −, +, −,....) are random.



MODULE QUIZ 8.2

1. Which of the following assumptions is *least likely* required for the difference in means test based on two samples?
   A. The two samples are independent.
   B. The two populations are normally distributed.
   C. The two populations have known variances.
2. The appropriate test statistic for a test of the equality of variances for two normally distributed random variables, based on two independent random samples, is the:
   A. *t*-test.
   B. *F*-test.
   C. χ^2 test.
3. The appropriate test statistic to test the hypothesis that the variance of a normally distributed population is equal to 13 is the:
   A. *t*-test.
   B. *F*-test.
   C. χ^2 test.

KEY CONCEPTS

LOS 8.a

The hypothesis testing process requires a statement of a null and an alternative hypothesis, the selection of the appropriate test statistic, specification of the significance level, a decision rule, the calculation of a sample statistic, a decision regarding the hypotheses based on the test, and a decision based on the test results.

The null hypothesis is what the researcher wants to reject. The alternative hypothesis is what the researcher wants to support, and it is accepted when the null hypothesis is rejected.

A Type I error is the rejection of the null hypothesis when it is actually true, while a Type II error is the failure to reject the null hypothesis when it is actually false.

The significance level can be interpreted as the probability that a test statistic will reject the null hypothesis by chance when it is actually true (i.e., the probability of a Type I error). A significance level must be specified to select the critical values for the test.

The power of a test is the probability of rejecting the null when it is false. The power of a test = 1 − P(Type II error).

The *p*-value for a hypothesis test is the smallest significance level for which the hypothesis would be rejected.

LOS 8.b

Hypothesis tests of:	Use a:	With degrees of freedom:
One population mean	*t*-statistic	$n - 1$
Two population means	*t*-statistic	$n - 1$
One population variance	Chi-square statistic	$n - 1$
Two population variances	*F*-statistic	$n_1 - 1, n_2 - 1$

LOS 8.c

Parametric tests, like the *t*-test, *F*-test, and chi-square test, make assumptions regarding the distribution of the population from which samples are drawn. Nonparametric tests either do not consider a particular population parameter or have few assumptions about the sampled population. Nonparametric tests are used when the assumptions of parametric tests can't be supported, or when the data are not suitable for parametric tests.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 8.1

1. **C** A Type I error is rejecting the null hypothesis when it is true. The probability of rejecting a false null is [1 − Prob Type II] = [1 − 0.60] = 40%, which is called the power of the test. The other answer choices are not necessarily true, because the null may be false and the probability of rejection unknown. (LOS 8.a)

2. **A** The power of a test is 1 − P(Type II error) = 1 − 0.15 = 0.85. (LOS 8.a)

Module Quiz 8.2

1. **C** The difference in means test does not require the two population variances to be known. (LOS 8.b)

2. **B** The *F*-test is the appropriate test. (LOS 8.b)

3. **C** A test of the population variance is a chi-square test. (LOS 8.b)

READING 9

PARAMETRIC AND NON-PARAMETRIC TESTS OF INDEPENDENCE

MODULE 9.1: TESTS FOR INDEPENDENCE

Video covering this content is available online.

LOS 9.a: Explain parametric and nonparametric tests of the hypothesis that the population correlation coefficient equals zero, and determine whether the hypothesis is rejected at a given level of significance.

Correlation measures the strength of the relationship between two variables. If the correlation between two variables is zero, there is no linear relationship between them. When the sample correlation coefficient for two variables is different from zero, we must address the question of whether the true population correlation coefficient (ρ) is equal to zero. The appropriate test statistic for the hypothesis that the population correlation equals zero, when the two variables are normally distributed, is as follows:

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

where:
r = sample correlation
n = sample size

This test statistic follows a *t*-distribution with $n - 2$ degrees of freedom. Note that the test statistic increases, not only with the sample correlation coefficient, but also with sample size.

EXAMPLE: Test of the hypothesis that the population correlation coefficient equals zero

A researcher computes the sample correlation coefficient for two normally distributed random variables as 0.35, based on a sample size of 42. Determine whether to reject the hypothesis that the population correlation coefficient is equal to zero at a 5% significance level.

Answer:

Our test statistic is $\frac{0.35\sqrt{42-2}}{\sqrt{1-0.35^2}} = 2.363$.

Using the *t*-table with 42 − 2 = 40 degrees of freedom for a two-tailed test and a significance level of 5%, we can find the critical value of 2.021. Because our computed test statistic of 2.363 is greater than 2.021, we reject the hypothesis that the population mean is zero and conclude that it is not equal to zero. That is, the two populations are correlated—in this case, positively.



PROFESSOR'S NOTE

The correlation coefficient we refer to here is the Pearson correlation coefficient, which is a measure of the linear relationship between two variables. There are other correlation coefficients that better measure the strength of any nonlinear relationship between two variables.

The **Spearman rank correlation test**, a nonparametric test, can be used to test whether two sets of ranks are correlated. Ranks are simply ordered values. If there is a tie (equal values), the ranks are shared—so if second and third rank is the same, the ranks are shared, and each gets a rank of (2 + 3) / 2 = 2.5.

The Spearman rank correlation, r_s (when all ranks are integer values), is calculated as follows:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

where:

r_S = rank correlation

n = sample size

d_i = difference between two ranks

We can test the significance of the Spearman rank correlation calculated with the formula just listed using the same test statistic we used for estimating the significance of a parametric correlation coefficient:

$$\frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}}$$

When the sample size is greater than 30, the test statistic follows a *t*-distribution with n − 2 degrees of freedom.

LOS 9.b: Explain tests of independence based on contingency table data.

A contingency or two-way table shows the number of observations from a sample that have a combination of two characteristics. Figure 9.1 is a contingency table where the characteristics are earnings growth (low, medium, or high) and dividend yield (low, medium, or high). We can use the data in the table to test the hypothesis that the two characteristics, earnings growth and dividend yield, are independent of each other.

Figure 9.1: Contingency Table for Categorical Data

Earnings Growth	Dividend Yield			
	Low	Medium	High	Total
Low	28	53	42	123
Medium	42	32	39	113
High	49	25	14	88
Total	119	110	95	324

We index our three categories of earnings growth from low to high with i = 1, 2, or 3, and our three categories of dividend yield from low to high with j = 1, 2, or 3. From the table, we see in Cell 1,1 that 28 firms have both low earnings growth and low dividend yield. We see in Cell 3,2 that 25 firms have high earnings growth and medium dividend yields.

For our test, we are going to compare the actual table values to what the values would be if the two characteristics were independent. The test statistic is a chi-square test statistic calculated as follows:

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

where:

O_{ij} = the number of observations in Cell i,j: Row i and Column j (i.e., observed frequency)

E_{ij} = the expected number of observations for Cell i,j

r = the number of row categories

c = the number of column categories

The degrees of freedom are (r − 1) × (c − 1), which is 4 in our example for dividend yield and earnings growth.

E_{ij}, the expected number of observations in Cell i,j, is:

$$\frac{\text{total for Row i} \times \text{total for Column j}}{\text{total for all columns and rows}}$$

The expected number of observations for Cell 2,2 is:

$$\frac{110 \times 113}{324} = 38.4$$

In calculating our test statistic, the term for Cell 2,2 is:

$$\frac{(32-38.4)^2}{38.4} = 1.0667$$

Figure 9.2 shows the expected frequencies for each pair of categories in our earnings growth and dividend yield contingency table.

Figure 9.2: Contingency Table for Expected Frequencies

Earnings Growth	Dividend Yield		
	Low	Medium	High
Low	45.2	41.8	36.1
Medium	41.5	38.4	33.1
High	32.3	29.9	25.8

For our test statistic, we sum, for all nine cells, the squared difference between the expected frequency and observed frequency, divided by the expected frequency. The resulting sum is 27.47. Figure 9.3 shows the results for each cell in calculating the test statistic.

Figure 9.3: Squared Differences from Contingency Table

Earnings Growth	Dividend Yield		
	Low	Medium	High
Low	6.5451	3.0010	0.9643
Medium	0.0060	1.0667	1.0517
High	8.6344	0.8030	5.3969
	Sum = 27.4691		

Our degrees of freedom are (3 − 1) × (3 − 1) = 4. The critical value for a significance level of 5% (from the chi-square table in the Appendix) with 4 degrees of freedom is 9.488. Based on our sample data, we can reject the hypothesis that the earnings growth and dividend yield categories are independent.



MODULE QUIZ 9.1

1. The test statistic for a Spearman rank correlation test for a sample size greater than 30 follows a:
   A. *t*-distribution.
   B. normal distribution.
   C. chi-square distribution.
2. A contingency table can be used to test:
   A. a null hypothesis that rank correlations are equal to zero.
   B. whether multiple characteristics of a population are independent.
   C. the number of *p*-values from multiple tests that are less than adjusted critical values.
3. For a parametric test of whether a correlation coefficient is equal to zero, it is *least likely* that:
   A. degrees of freedom are n − 1.
   B. the test statistic follows a *t*-distribution.
   C. the test statistic increases with a greater sample size.

KEY CONCEPTS

LOS 9.a

To test a hypothesis that a population correlation coefficient equals zero, the appropriate test statistic is a *t*-statistic with n − 2 degrees of freedom, calculated as $\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$, where *r* is the sample correlation coefficient.

A nonparametric test of correlation can be performed when we have only ranks (e.g., deciles of investment performance). The Spearman rank correlation test examines whether the ranks for multiple periods are correlated. The rank correlation is $r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$, where d_i^2 is the sum of the squared differences in pairs of ranks and *n* is the number of sample periods. The test statistic follows a *t*-distribution for samples sizes greater than 30.

LOS 9.b

A contingency table can be used to test the hypothesis that two characteristics (categories) of a sample of items are independent. A contingency table shows the number of the sample items (e.g., firms that have both of two characteristics). The test statistic follows a chi-square distribution and is calculated as follows:

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

where:

O_{ij} = the number of observations in Cell i,j: Row *i* and Column *j* (i.e., observed frequency)

E_{ij} = the expected number of observations for Cell i,j of the contingency table with independence

r = the number of row categories

c = the number of column categories

The degrees of freedom are (r − 1) × (c − 1). If the test statistic is greater than the critical chi-square value for a given level of significance, we reject the hypothesis that the two characteristics are independent.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 9.1

1. **A** The test statistic for a Spearman rank correlation test follows a *t*-distribution. (LOS 9.a)

2. **B** A contingency table is used to determine whether two characteristics of a group are independent. (LOS 9.b)

3. **A** Degrees of freedom are n − 2 for a test of the hypothesis that correlation is equal to zero. The test statistic increases with sample size (degrees of freedom increase) and follows a *t*-distribution. (LOS 9.a)

READING 10

SIMPLE LINEAR REGRESSION

MODULE 10.1: LINEAR REGRESSION BASICS

Video covering this content is available online.

LOS 10.a: Describe a simple linear regression model, how the least squares criterion is used to estimate regression coefficients, and the interpretation of these coefficients.

The purpose of **simple linear regression** is to explain the variation in a dependent variable in terms of the variation in a single independent variable. Here, the term *variation* is interpreted as the degree to which a variable differs from its mean value. Don't confuse *variation* with *variance*—they are related, but they are not the same.

$$\text{variation in Y} = \sum_{i}^{n}(Y_i - \overline{Y})^2$$

- The **dependent variable** is the variable whose variation is explained by the independent variable. We are interested in answering the question, "What explains fluctuations in the dependent variable?" The dependent variable is also referred to as the terms *explained variable, endogenous variable,* or *predicted variable*.
- The **independent variable** is the variable used to explain the variation of the dependent variable. The independent variable is also referred to as the terms *explanatory variable, exogenous variable,* or *predicting variable*.

EXAMPLE: Dependent vs. independent variables

Suppose you want to predict stock returns with GDP growth. Which variable is the independent variable?

Answer:

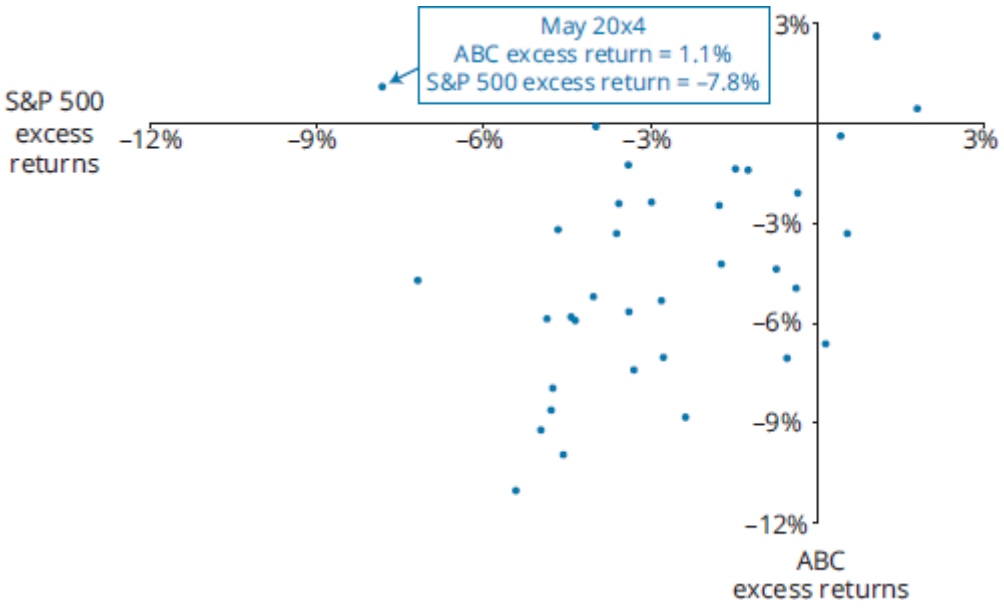
Because GDP is going to be used as a predictor of stock returns, stock returns are being *explained* by GDP. Hence, stock returns are the dependent (explained) variable, and GDP is the independent (explanatory) variable.

Suppose we want to use excess returns on the S&P 500 (the independent variable) to explain the variation in excess returns on ABC common stock (the dependent variable).

For this model, we define excess return as the difference between the actual return and the return on 1-month Treasury bills.

We would start by creating a scatter plot with ABC excess returns on the vertical axis and S&P 500 excess returns on the horizontal axis. Monthly excess returns for both variables from June 20X2 to May 20X5 are plotted in Figure 10.1. For example, look at the point labeled May 20X4. In that month, the excess return on the S&P 500 was −7.8%, and the excess return on ABC was 1.1%.

Figure 10.1: Scatter Plot of ABC Excess Returns vs. S&P 500 Index Excess Returns



The two variables in Figure 10.1 appear to be positively correlated: excess ABC returns tended to be positive (negative) in the same month that S&P 500 excess returns were positive (negative). This is not the case for all the observations, however (for example, May 20X4). In fact, the correlation between these variables is approximately 0.40.

Simple Linear Regression Model

The following linear regression model is used to describe the relationship between two variables, *X* and *Y*:

$$Y_i = b_0 + b_1 X_i + \varepsilon_i, \ldots i = 1, \ldots, n$$

where:

Y_i = *i*th observation of the dependent variable, *Y*
X_i = *i*th observation of the independent variable, *X*
b_0 = regression intercept term
b_1 = regression slope coefficient
ε_i = **residual** for the *i*th observation (also referred to as the disturbance term or error term)

Based on this regression model, the regression process estimates an equation for a line through a scatter plot of the data that "best" explains the observed values for *Y* in terms of the observed values for *X*.

The linear equation, often called the line of best fit or **regression line**, takes the following form:

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i, i = 1,2,3...,n$$

where:

$\hat{Y}_i$ = estimated value of Y_i given X_i

$\hat{b}_0$ = estimated intercept term

$\hat{b}_1$ = estimated slope coefficient



PROFESSOR'S NOTE

The hat "^" above a variable or parameter indicates a predicted value.

The regression line is just one of the many possible lines that can be drawn through the scatter plot of *X* and *Y*. The criteria used to estimate this line is the essence of linear regression. The regression line is the line that minimizes the sum of the squared differences (vertical distances) between the *Y*-values predicted by the regression equation ($\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i$) and the actual *Y*-values, Y_i. The sum of the squared vertical distances between the estimated and actual *Y*-values is referred to as the **sum of squared errors (SSE)**.

Thus, the regression line is the line that minimizes the SSE. This explains why simple linear regression is frequently referred to as **ordinary least squares (OLS)** regression, and the values determined by the estimated regression equation, $\hat{Y}_i$, are called least squares estimates.

The estimated **slope coefficient** ($\hat{b}_1$) for the regression line describes the change in *Y* for a one-unit change in *X*. It can be positive, negative, or zero, depending on the relationship between the regression variables. The slope term is calculated as follows:

$$\hat{b}_1 = \frac{Cov_{XY}}{\sigma_X^2}$$

The intercept term ($\hat{b}_0$) is the line's intersection with the *Y*-axis at X = 0. It can be positive, negative, or zero. A property of the least squares method is that the intercept term may be expressed as follows:

$$\hat{b}_0 = \overline{Y} - \hat{b}_1 \overline{X}$$

where:

$\overline{Y}$ = mean of Y

$\overline{X}$ = mean of X

The intercept equation highlights the fact that the regression line passes through a point with coordinates equal to the mean of the independent and dependent variables (i.e., the point X, Y).

EXAMPLE: Computing the slope coefficient and intercept term

Compute the slope coefficient and intercept term using the following information:

Cov(S&P 500, ABC) = 0.000336 Mean return, S&P 500 = −2.70%

Var(S&P 500) = 0.000522 Mean return, ABC = −4.05%

Answer:

The slope coefficient is calculated as $\hat{b}_1 = 0.000336 / 0.000522 = 0.64$.

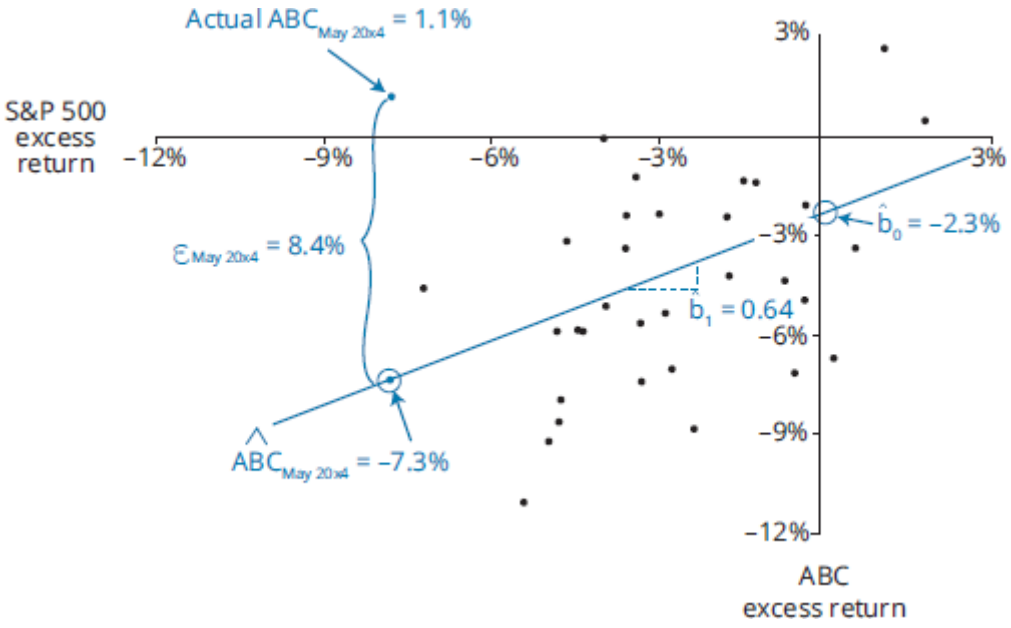
The intercept term is calculated as follows:

$$\hat{b}_0 = \overline{ABC} - \hat{b}_1 \overline{S\&P\ 500} = -4.05\% - 0.64(-2.70\%) = -2.3\%$$

The estimated regression line that minimizes the SSE in our ABC stock return example is shown in Figure 10.2.

This regression line has an intercept of −2.3% and a slope of 0.64. The model predicts that if the S&P 500 excess return is −7.8% (May 20X4 value), then the ABC excess return would be −2.3% + (0.64)(−7.8%) = −7.3%. The residual (error) for the May 20X4 ABC prediction is 8.4%—the difference between the actual ABC excess return of 1.1% and the predicted return of −7.3%.

Figure 10.2: Estimated Regression Equation for ABC vs. S&P 500 Excess Returns



Interpreting a Regression Coefficient

The estimated intercept represents the value of the dependent variable at the point of intersection of the regression line and the axis of the dependent variable (usually, the vertical axis). In other words, the intercept is an estimate of the dependent variable when the independent variable is zero.

We also mentioned earlier that the estimated slope coefficient is interpreted as the expected change in the dependent variable for a one-unit change in the independent variable. For example, an estimated slope coefficient of 2 would indicate that the dependent variable is expected to change by two units for every one-unit change in the independent variable.

EXAMPLE: Interpreting regression coefficients

In the previous example, the estimated slope coefficient was 0.64 and the estimated intercept term was −2.3%. Interpret each coefficient estimate.

Answer:

The slope coefficient of 0.64 can be interpreted to mean that when excess S&P 500 returns increase (decrease) by 1%, ABC excess returns is expected to increase (decrease) by 0.64%.

The intercept term of −2.3% can be interpreted to mean that when the excess return on the S&P 500 is zero, the expected return on ABC stock is −2.3%.



PROFESSOR'S NOTE

The slope coefficient in a regression of the excess returns of an individual security (the *y*-variable) on the return on the market (the *x*-variable) is called the stock's beta, which is an estimate of systematic risk of ABC stock. Notice that ABC is less risky than the average stock, because its returns tend to increase or decrease by less than the overall change in the market returns. A stock with a beta (regression slope coefficient) of 1 has an average level of systematic risk, and a stock with a beta greater than 1 has more-than-average systematic risk. We will apply this concept in the Portfolio Management topic area.

Keep in mind, however, that any conclusions regarding the importance of an independent variable in explaining a dependent variable are based on the statistical significance of the slope coefficient. The magnitude of the slope coefficient tells us nothing about the strength of the linear relationship between the dependent and independent variables. A hypothesis test must be conducted, or a confidence interval must be formed, to assess the explanatory power of the independent variable. Later in this reading we will perform these hypothesis tests.

LOS 10.b: Explain the assumptions underlying the simple linear regression model, and describe how residuals and residual plots indicate if these assumptions may have been violated.

Linear regression is based on numerous assumptions. Most of the major assumptions pertain to the regression model's residual term (ε). Linear regression assumes the following:

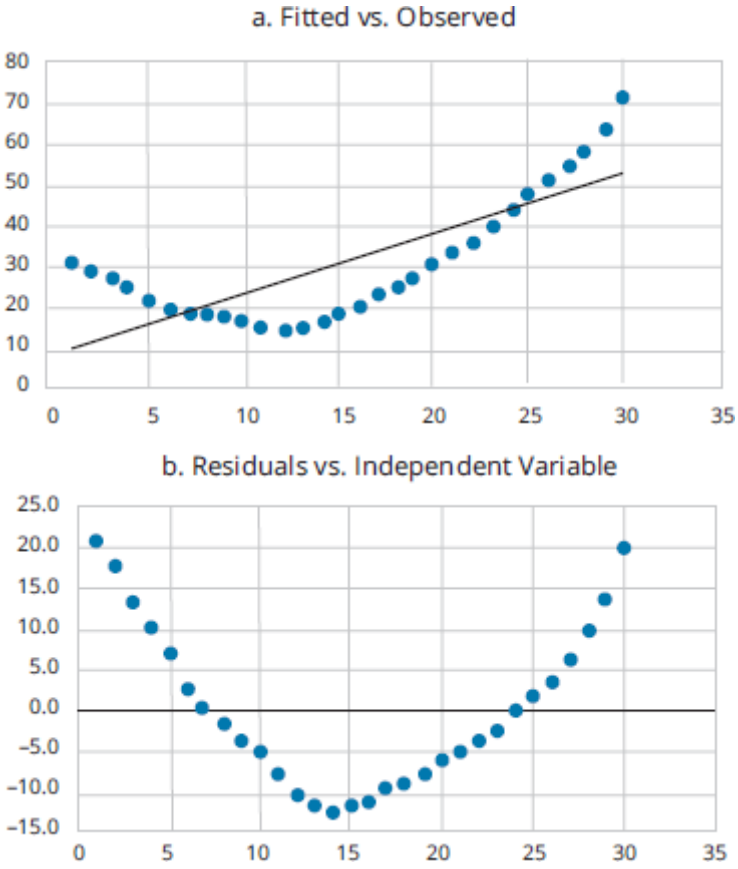
1. A linear relationship exists between the dependent and the independent variables.
2. The variance of the residual term is constant for all observations (homoskedasticity).
3. The residual term is independently distributed; that is, the residual for one observation is not correlated with that of another observation (or, the paired *x* and *y* observations are independent of each other).

4. The residual term is normally distributed.

Linear Relationship

A linear regression model is not appropriate when the underlying relationship between *X* and *Y* is nonlinear. In Panel A of Figure 10.3, we illustrate a regression line fitted to a nonlinear relationship. Note that the prediction errors (vertical distances from the dots to the line) are positive for low values of *X*, then increasingly negative for higher values of *X*, and then turning positive for still-greater values of *X*. One way of checking for linearity is to examine the model residuals (prediction errors) in relation to the independent regression variable. In Panel B, we show the pattern of residuals over the range of the independent variable: positive, negative, then positive.

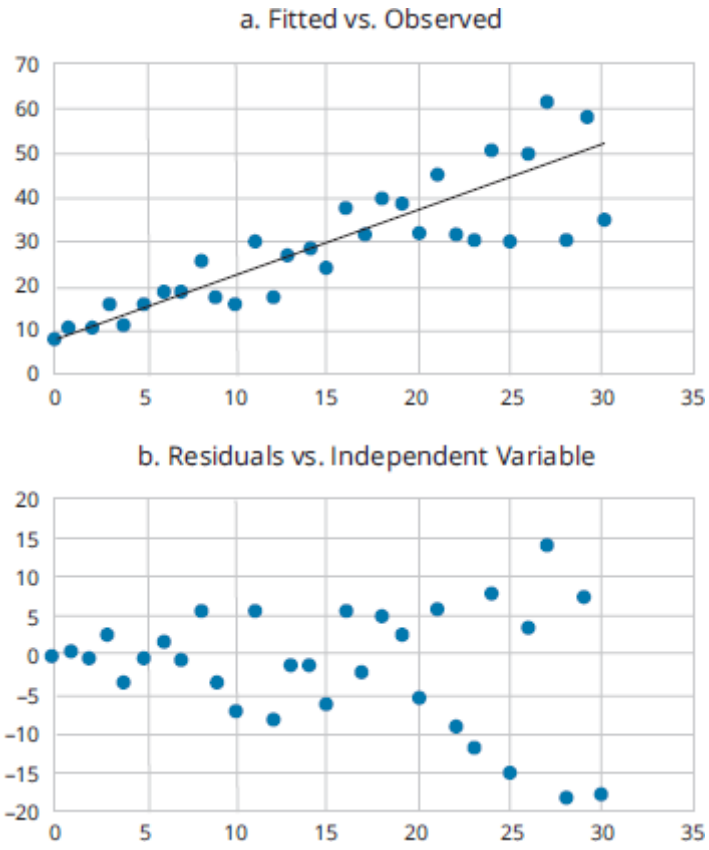
Figure 10.3: Nonlinear Relationship



Homoskedasticity

Homoskedasticity refers to the case where prediction errors all have the same variance. **Heteroskedasticity** refers to the situation when the assumption of homoskedasticity is violated. Figure 10.4, Panel A shows a scatter plot of observations around a fitted regression line where the residuals (prediction errors) increase in magnitude with larger values of the independent variable *X*. Panel B shows the residuals plotted versus the value of the independent variable, and it also illustrates that the variance of the error terms is not likely constant for all observations.

Figure 10.4: Heteroskedasticity

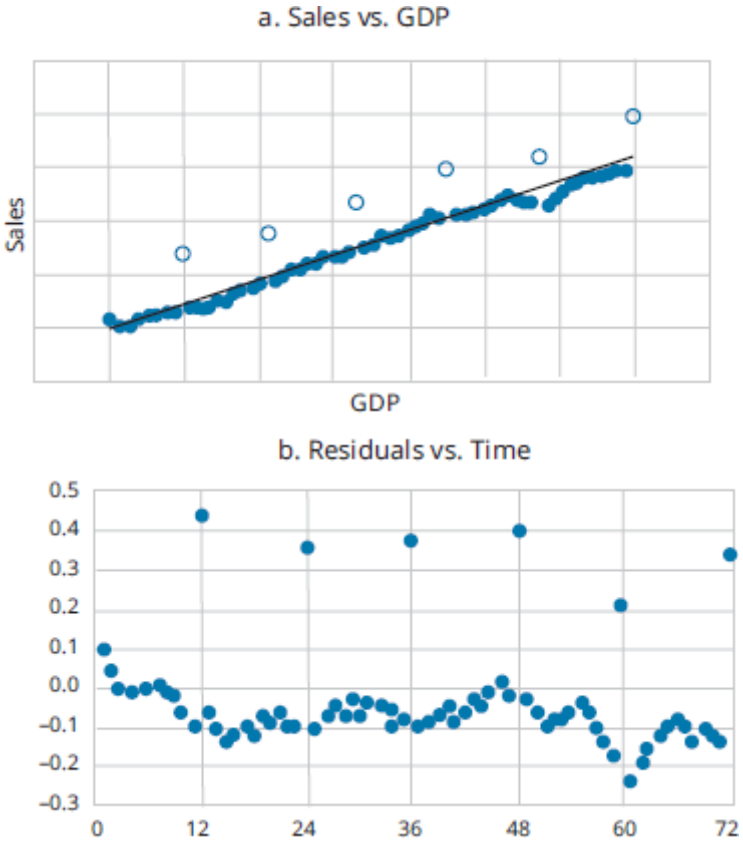


Another type of heteroskedasticity results if the variance of the error term changes over time (rather than with the magnitude of the independent variable). We could observe this by plotting the residuals from a linear regression model versus the dates of each observation and finding that the magnitude of the errors exhibits a pattern of changing over time. To illustrate this, we could plot the residuals versus a time index (as the *x*-variable). Residuals would exhibit a pattern of increasing over time.

Independence

Suppose we collect a company's monthly sales and plot them against monthly GDP as in Figure 10.5, Panel A, and observe that some prediction errors (the unfilled dots) are noticeably larger than others. To investigate this, we plot the residuals versus time, as in Panel B. The residuals plot illustrates that there are large negative prediction errors every 12 months (in December). This suggests that there is seasonality in sales such that December sales (the unfilled dots in Figure 10.5) are noticeably farther from their predicted values than sales for the other months. If the relationship between *X* and *Y* is not independent, the residuals are not independent, and our estimates of the model parameters' variances will not be correct.

Figure 10.5: Independence



Normality

When the residuals (prediction errors) are normally distributed, we can conduct hypothesis testing for evaluating the goodness of fit of the model (discussed later). With a large sample size, based on the central limit theorem, our parameter estimates may be valid, even when the residuals are not normally distributed.

Outliers are observations (one or a few) that are far from our regression line (have large prediction errors or *X* values that are far from the others). Outliers will influence our parameter estimates so that the OLS model will not fit the other observations well.

MODULE QUIZ 10.1

1. What is the *most appropriate* interpretation of a slope coefficient estimate equal to 10.0?
   A. The predicted value of the dependent variable when the independent variable is 0 is 10.0.
   B. For every 1-unit change in the independent variable, the model predicts that the dependent variable will change by 10 units.
   C. For every 1-unit change in the independent variable, the model predicts that the dependent variable will change by 0.1 units.
2. Which of the following is *least likely* a necessary assumption of simple linear regression analysis?
   A. The residuals are normally distributed.
   B. There is a constant variance of the error term.

C. The dependent variable is uncorrelated with the residuals.

MODULE 10.2: ANALYSIS OF VARIANCE (ANOVA) AND GOODNESS OF FIT



Video covering this content is available online.

LOS 10.c: Calculate and interpret measures of fit and formulate and evaluate tests of fit and of regression coefficients in a simple linear regression.

LOS 10.d: Describe the use of analysis of variance (ANOVA) in regression analysis, interpret ANOVA results, and calculate and interpret the standard error of estimate in a simple linear regression.

Analysis of variance (ANOVA) is a statistical procedure for analyzing the total variability of the dependent variable. Let's define some terms before we move on to ANOVA tables:

- The **total sum of squares (SST)** measures the total variation in the dependent variable. SST is equal to the sum of the squared differences between the actual *Y*-values and the mean of *Y*:

$$\text{SST} = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

- The **sum of squares regression (SSR)** measures the variation in the dependent variable that is explained by the independent variable. SSR is the sum of the squared distances between the predicted *Y*-values and the mean of *Y*:

$$\text{SSR} = \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2$$

- The **mean square regression (MSR)** is the SSR divided by the number of independent variables. A simple linear regression has only one independent variable, so in this case, MSR = SSR.



PROFESSOR'S NOTE
Multiple regression (i.e., with more than one independent variable) is addressed in the Level II CFA curriculum.

- The **sum of squared errors (SSE)** measures the unexplained variation in the dependent variable. It's also known as the sum of squared residuals or the residual sum of squares. SSE is the sum of the squared vertical distances between the actual *Y*-values and the predicted *Y*-values on the regression line:

$$\text{SSE} = \sum_{i=1}^{n}(Y_i - \hat{Y})^2$$

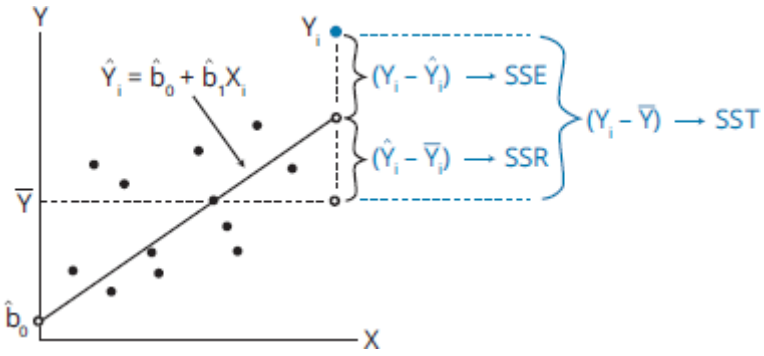
- The **mean squared error (MSE)** is the SSE divided by the degrees of freedom, which is $n - 1$ minus the number of independent variables. A simple linear regression has only one independent variable, so in this case, degrees of freedom are $n - 2$.

You probably will not be surprised to learn the following:

$$\text{total variation} = \text{explained variation} + \text{unexplained variation}$$
or:
$$\text{SST} = \text{SSR} + \text{SSE}$$

Figure 10.6 illustrates how the total variation in the dependent variable (SST) is composed of SSR and SSE.

Figure 10.6: Components of Total Variation



The output of the ANOVA procedure is an ANOVA table, which is a summary of the variation in the dependent variable. ANOVA tables are included in the regression output of many statistical software packages. You can think of the ANOVA table as the source of the data for the computation of many of the regression concepts discussed in this reading. A generic ANOVA table for a simple linear regression (one independent variable) is presented in Figure 10.7.

Figure 10.7: ANOVA Table for a Simple Linear Regression

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Sum of Squares
Regression (explained)	1	SSR	$\text{MSR} = \frac{\text{SSR}}{k} = \frac{\text{SSR}}{1} = \text{SSR}$
Error (unexplained)	$n - 2$	SSE	$\text{MSE} = \frac{\text{SSE}}{n-2}$
Total	$n - 1$	SST	

Standard Error of Estimate (SEE)

The SEE for a regression is the standard deviation of its residuals. The lower the SEE, the better the model fit:

$$\text{SEE} = \sqrt{\text{MSE}}$$

Coefficient of Determination (R^2)

The **coefficient of determination** (R^2) is defined as the percentage of the total variation in the dependent variable explained by the independent variable. For example, an R^2 of 0.63 indicates that the variation of the independent variable explains 63% of the variation in the dependent variable:

$$R^2 = \text{SSR} / \text{SST}$$

PROFESSOR'S NOTE
For simple linear regression (i.e., with one independent variable), the coefficient of determination, R^2, may be computed by simply squaring the correlation coefficient, *r*. In other words, $R^2 = r^2$ for a regression with one independent variable.

EXAMPLE: Using the ANOVA table

Given the following ANOVA table based on 36 observations, calculate the R^2 and the standard error of estimate (SEE).

Completed ANOVA table for ABC regression

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Sum of Squares
Regression (explained)	1	0.0076	0.0076
Error (unexplained)	34	0.0406	0.0012
Total	35	0.0482	

Answer:

$$R^2 = \frac{\text{explained variation (SSR)}}{\text{total variation (SST)}} = \frac{0.0076}{0.0482} = 0.158 \text{ or } 15.8\%$$

$$\text{SEE} = \sqrt{\text{MSE}} = \sqrt{0.0012} = 0.035$$

The *F*-Statistic

An *F*-test assesses how well a set of independent variables, as a group, explains the variation in the dependent variable.

The *F*-statistic is calculated as follows:

$$F = \frac{\text{MSR}}{\text{MSE}} = \frac{\text{SSR}/k}{\text{SSE}/n - k - 1}$$

where:

MSR = mean regression sum of squares
MSE = mean squared error
Important: This is always a one-tailed test!

For simple linear regression, there is only one independent variable, so the *F*-test is equivalent to a *t*-test of the statistical significance of the slope coefficient:

$$H_0: b_1 = 0 \text{ versus } H_a: b_1 \neq 0$$

To determine whether b_1 is statistically significant using the *F*-test, the calculated *F*-statistic is compared with the critical *F*-value, F_c, at the appropriate level of significance. The degrees of freedom for the numerator and denominator with one independent variable are as follows:

$$df_{numerator} = k = 1$$
$$df_{denominator} = n - k - 1 = n - 2$$

where:

n = number of observations

The decision rule for the *F*-test is to reject H_0 if $F > F_c$.

Rejecting the null hypothesis that the value of the slope coefficient equals zero at a stated level of significance indicates that the independent variable and the dependent variable have a significant linear relationship.

EXAMPLE: Calculating and interpreting the *F*-statistic

Use the ANOVA table from the previous example to calculate and interpret the *F*-statistic. Test the null hypothesis at the 5% significance level that the slope coefficient is equal to 0.

Answer:

$$F = \frac{MSR}{MSE} = \frac{0.0076}{0.0012} = 6.33$$
$$df_{numerator} = k = 1$$
$$df_{denominator} = n - k - 1 = 36 - 1 - 1 = 34$$

The null and alternative hypotheses are h_0: $b_1 = 0$ versus h_a: $b_1 \neq 0$. The critical *F*-value for 1 and 34 degrees of freedom at a 5% significance level is approximately 4.1. (Remember, it's a one-tailed test, so we use the 5% *F*-table.) Therefore, we can reject the null hypothesis and conclude that the slope coefficient is significantly different than zero.

Hypothesis Test of a Regression Coefficient

A *t*-test may also be used to test the hypothesis that the true slope coefficient, b_1, is equal to a hypothesized value. Letting $\hat{b}_1$ be the point estimate for b_1, the appropriate test statistic with n – 2 degrees of freedom is:

$$t_{b_1} = \frac{\hat{b}_1 - b_1}{s_{\hat{b}_1}}$$

The decision rule for tests of significance for regression coefficients is:

$$\text{Reject } H_0 \text{ if } t > +t_{critical} \text{ or } t < -t_{critical}$$

Rejection of the null supports the alternative hypothesis that the slope coefficient is *different* from the hypothesized value of b_1. To test whether an independent variable explains the variation in the dependent variable (i.e., it is statistically significant), the null hypothesis is that the true slope is zero ($b_1 = 0$). The appropriate test structure for the null and alternative hypotheses is:

$$H_0: b_1 = 0 \text{ versus } H_a: b_1 \neq 0$$

EXAMPLE: Hypothesis test for significance of regression coefficients

The estimated slope coefficient from the ABC example is 0.64 with a standard error equal to 0.26. Assuming that the sample has 36 observations, determine if the estimated slope coefficient is significantly different than zero at a 5% level of significance.

Answer:

The calculated test statistic is:

$$t = \frac{\hat{b}_1 - b_1}{s_{\hat{b}_1}} = \frac{0.64 - 0}{0.26} = 2.46$$

The critical two-tailed *t*-values are ±2.03 (from the *t*-table with df = 36 − 2 = 34). Because $t > t_{critical}$ (i.e., 2.46 > 2.03), we reject the null hypothesis and conclude that the slope is different from zero.

Note that the *t*-test for a simple linear regression is equivalent to a *t*-test for the correlation coefficient between *x* and *y*:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$



MODULE QUIZ 10.2

1. Consider the following statement: "In a simple linear regression, the appropriate degrees of freedom for the critical *t*-value used to calculate a confidence interval around both a parameter estimate and a predicted *Y*-value is the same as the number of observations minus two." This statement is:
   A. justified.
   B. not justified, because the appropriate degrees of freedom used to calculate a confidence interval around a parameter estimate is the number of observations.
   C. not justified, because the appropriate degrees of freedom used to calculate a confidence interval around a predicted *Y*-value is the number of observations.

2. What is the appropriate alternative hypothesis to test the statistical significance of the intercept term in the following regression?

   $Y = a_1 + a_2(X) + \varepsilon$

   A. $H_A: a_1 \neq 0$.
   B. $H_A: a_1 > 0$.
   C. $H_A: a_2 \neq 0$.

3. The variation in the dependent variable explained by the independent variable is measured by the:
   A. mean squared error.
   B. sum of squared errors.

C. regression sum of squares.

MODULE 10.3: PREDICTED VALUES AND FUNCTIONAL FORMS OF REGRESSION



Video covering this content is available online.

LOS 10.e: Calculate and interpret the predicted value for the dependent variable, and a prediction interval for it, given an estimated linear regression model and a value for the independent variable.

Predicted values are values of the dependent variable based on the estimated regression coefficients and a prediction about the value of the independent variable. They are the values that are *predicted* by the regression equation, given an estimate of the independent variable.

For a simple regression, this is the predicted (or forecast) value of *Y*:

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X_p$$

where:

$\hat{Y}$ = predicted value of the dependent variable
X_p = forecasted value of the independent variable

EXAMPLE: Predicting the dependent variable

Given the ABC regression equation as follows:

$$\widehat{ABC} = -2.3\% + (0.64)(\widehat{S\&P\ 500})$$

Calculate the predicted value of ABC excess returns if forecast S&P 500 excess returns are 10%.

Answer:

The predicted value for ABC excess returns is determined as follows:

$$\widehat{ABC} = -2.3\% + (0.64)(10\%) = 4.1\%$$

Confidence Intervals for Predicted Values

This is the equation for the confidence interval for a predicted value of *Y*:

$$\hat{Y} \pm (t_c \times s_f) \Rightarrow [\hat{Y} - (t_c \times s_f) < Y < \hat{Y} + (t_c \times s_f)]$$

where:

t_c = two-tailed critical *t*-value at the desired level of significance with df = n − 2
s_f = standard error of the forecast

The challenge with computing a confidence interval for a predicted value is calculating s_f. On the Level I exam, it's highly unlikely that you will have to calculate the standard

error of the forecast (it will probably be provided if you need to compute a confidence interval for the dependent variable). However, if you do need to calculate s_f, it can be done with the following formula for the variance of the forecast:

$$s_f^2 = SEE^2\left[1 + \frac{1}{n} + \frac{(X - \overline{X})^2}{(n-1)s_x^2}\right]$$

where:

SEE^2 = variance of the residuals = the square of the standard error of estimate

s_x^2 = variance of the independent variable

X = value of the independent variable for which the forecast was made

EXAMPLE: Confidence interval for a predicted value

Calculate a 95% prediction interval on the predicted value of ABC excess returns from the previous example. Assume the standard error of the forecast is 3.67, and the forecast value of S&P 500 excess returns is 10%.

Answer:

This is the predicted value for ABC excess returns:

$$\widehat{ABC} = -2.3\% + (0.64)(10\%) = 4.1\%$$

The 5% two-tailed critical *t*-value with 34 degrees of freedom is 2.03. This is the prediction interval at the 95% confidence level:

$$\widehat{ABC} \pm (t_c \times s_f) \Rightarrow [4.1\% \pm (2.03 \times 3.67\%)] = 4.1\% \pm 7.5\%$$

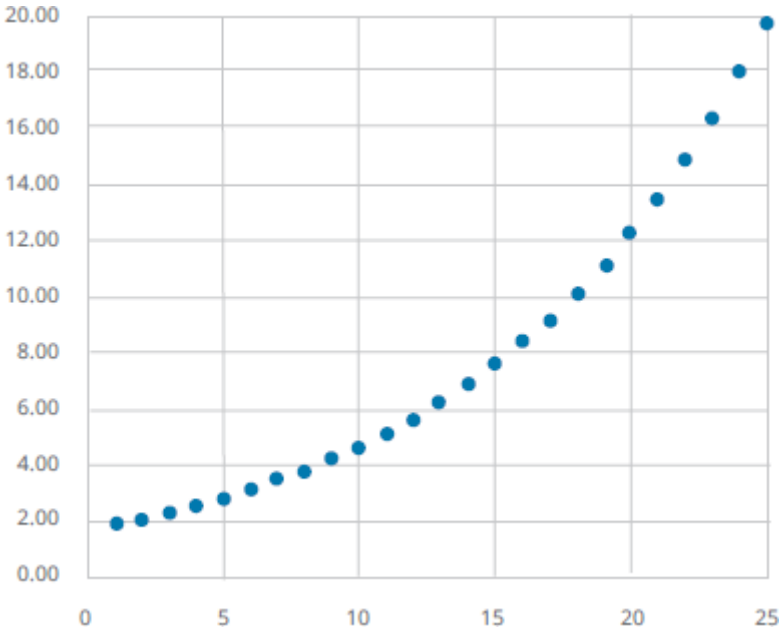
Or, −3.4% to 11.6%.

We can interpret this range to mean that, given a forecast value for S&P 500 excess returns of 10%, we can be 95% confident that the ABC excess returns will be between −3.4% and 11.6%.

LOS 10.f: Describe different functional forms of simple linear regressions.

One of the assumptions of linear regression is that the relationship between *X* and *Y* is linear. What if that assumption is violated? Consider *Y* = EPS for a company and *X* = time index. Suppose that EPS is growing at approximately 10% annually. Figure 10.8 shows the plot of actual EPS versus time.

Figure 10.8: Nonlinear Relationship



In such a situation, transforming one or both of the variables can produce a linear relationship. The appropriate transformation depends on the relationship between the two variables. One often-used transformation is to take the natural log of one or both of the variables. Here are some examples:

- **Log-lin model.** This is if the dependent variable is logarithmic, while the independent variable is linear.
- **Lin-log model.** This is if the dependent variable is linear, while the independent variable is logarithmic.
- **Log-log model.** Both the dependent variable and the independent variable are logarithmic.

Selecting the correct functional form involves determining the nature of the variables and evaluating the goodness-of-fit measures (e.g., R^2, SEE, *F*-stat).

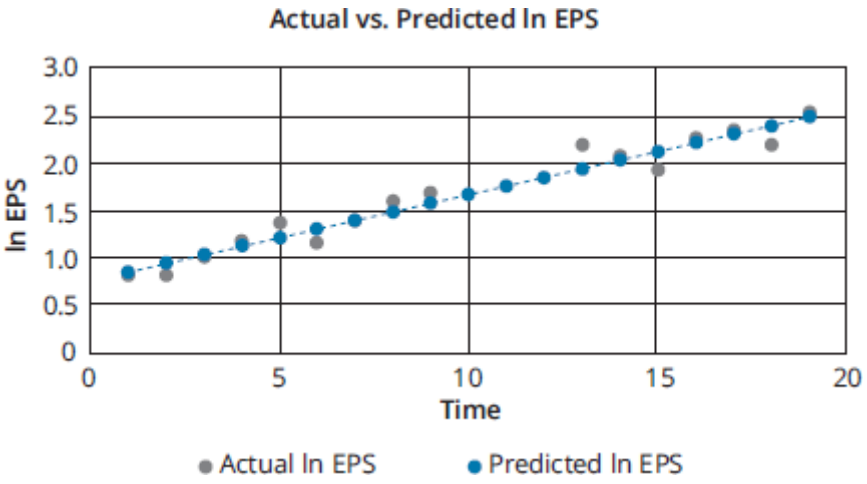
Log-Lin Model

Taking the natural logarithm of the dependent variable, our model now becomes this:

$$\ln Y_i = b_0 + b_1 X_i + \varepsilon_i$$

In this model, the slope coefficient is interpreted as the *relative* change in dependent variable for an absolute change in the independent variable. Figure 10.9 shows the results after taking the natural log of EPS, and fitting that data using a log-lin model.

Figure 10.9: Log-Lin Model, EPS Data



Lin-Log Model

Taking the natural logarithm of the independent variable, our model now becomes this:

$$Y_i = b_0 + b_1 \ln(X)_i + \varepsilon_i$$

In this model, the slope coefficient is interpreted as the *absolute* change in dependent variable for a *relative* change in the independent variable.

Log-Log Model

Taking the natural logarithm of both variables, our model now becomes this:

$$\ln Y_i = b_0 + b_1 \ln(X)_i + \varepsilon_i$$

In this model, the slope coefficient is interpreted as the relative change in dependent variable for a relative change in the independent variable.

MODULE QUIZ 10.3

1. For a regression model of Y = 5 + 3.5X, the analysis (based on a large data sample) provides the standard error of the forecast as 2.5 and the standard error of the slope coefficient as 0.8. A 90% confidence interval for the estimate of *Y* when the value of the independent variable is 10 is *closest* to:

   A. 35.1 to 44.9.
   B. 35.6 to 44.4.
   C. 35.9 to 44.1.

2. The appropriate regression model for a linear relationship between the relative change in an independent variable and the absolute change in the dependent variable is a:

   A. log-lin model.
   B. lin-log model.
   C. lin-lin model.

KEY CONCEPTS

LOS 10.a

Linear regression provides an estimate of the linear relationship between an independent variable (the explanatory variable) and a dependent variable (the predicted variable).

The general form of a simple linear regression model is as follows:

$$Y_i = b_0 + b_1 X_i + \varepsilon_i$$

The least squares model minimizes the sum of squared errors:

- $\hat{b}_0$ = fitted intercept = $\overline{Y} - \hat{b}_1 \overline{X}$
- $\hat{b}_1$ = fitted slope coefficient = Cov (X,Y) / variance of X

The estimated intercept, $\hat{b}_0$, represents the value of the dependent variable at the point of intersection of the regression line and the axis of the dependent variable (usually, the vertical axis). The estimated slope coefficient, $\hat{b}_1$, is interpreted as the change in the dependent variable for a one-unit change in the independent variable.

LOS 10.b

Assumptions made regarding simple linear regression include the following:

1. A linear relationship exists between the dependent and the independent variable.
2. The variance of the residual term is constant (homoskedasticity).
3. The residual term is independently distributed (residuals are uncorrelated).
4. The residual term is normally distributed.

LOS 10.c

The total sum of squares (SST) measures the total variation in the dependent variable and equals the sum of the squared differences between its actual values and its mean.

The sum of squares regression (SSR) measures the variation in the dependent variable that is explained by the independent variable, and equals the sum of the squared distances between the predicted values and the mean of the dependent variable.

The mean square regression (MSR) is the SSR divided by the number of independent variables. For a simple linear regression, MSR = SSR.

The sum of squared errors (SSE) measures the unexplained variation in the dependent variable and is the sum of the squared vertical distances between the actual and the predicted values of the dependent variable.

The mean squared error (MSE) is the SSE divided by the degrees of freedom (n − 2 for a simple linear regression).

The coefficient of determination, R^2, is the proportion of the total variation of the dependent variable explained by the regression:

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST}$$

In simple linear regression, because there is only one independent variable (k = 1), the *F*-test tests the same null hypothesis as testing the statistical significance of b_1 using

the *t*-test: H_0: $b_1 = 0$ versus H_a: $b_1 \neq 0$. With only one independent variable, *F* is calculated as follows:

$$\text{F-stat} = \frac{\text{MSR}}{\text{MSE}} \text{ with 1 and n} - 2 \text{ degrees of freedom}$$

LOS 10.d

ANOVA Table for Simple Linear Regression (k = 1)

Source of Variation	Degrees of Freedom (df)	Sum of Squares	Mean Sum of Squares
Regression (explained)	1	SSR	$\text{MSR} = \frac{\text{SSR}}{\text{k}} = \frac{\text{SSR}}{1} = \text{SSR}$
Error (unexplained)	n − 2	SSE	$\text{MSE} = \frac{\text{SSE}}{\text{n} - 2}$
Total	n − 1	SST	

The standard error of the estimate in a simple linear regression is calculated as follows:

$$\text{SEE} = \sqrt{\frac{\text{SSE}}{\text{n} - 2}}$$

LOS 10.e

A predicted value of the dependent variable, $\hat{Y}$, is determined by inserting the predicted value of the independent variable, X_p, in the regression equation and calculating $\hat{Y}_p = \hat{b}_0 + \hat{b}_1 \hat{X}_p$.

The confidence interval for a predicted *Y*-value is $\left[\hat{Y} - (t_c \times s_f) < Y < \hat{Y} + (t_c \times s_f)\right]$ where s_f is the standard error of the forecast.

LOS 10.f

Dependent Variable	Independent Variable	Model	Slope Interpretation
Logarithmic	Linear	Log-lin	*Relative* change in dependent variable for an *absolute* change in the independent variable
Linear	Logarithmic	Lin-log	*Absolute* change in dependent variable for a *relative* change in the independent variable
Logarithmic	Logarithmic	Log-log	*Relative* change in dependent variable for a *relative* change in the independent variable

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 10.1

1. **B** The slope coefficient is best interpreted as the predicted change in the dependent variable for a 1-unit change in the independent variable; if the slope coefficient estimate is 10.0 and the independent variable changes by 1 unit, the dependent variable is expected to change by 10 units. The intercept term is best interpreted

as the value of the dependent variable when the independent variable is equal to zero. (LOS 10.a)

2. **C** The model does not assume that the dependent variable is uncorrelated with the residuals. It does assume that the independent variable is uncorrelated with the residuals. (LOS 10.b)

Module Quiz 10.2

1. **A** In simple linear regression, the appropriate degrees of freedom for both confidence intervals is the number of observations in the sample (n) minus two. (LOS 10.c)

2. **A** In this regression, a_1 is the intercept term. To test the statistical significance means to test the null hypothesis that a_1 is equal to zero, versus the alternative that a_1 is not equal to zero. (LOS 10.c)

3. **C** The regression sum of squares measures the amount of variation in the dependent variable explained by the independent variable (i.e., the explained variation). The sum of squared errors measures the variation in the dependent variable not explained by the independent variable. The mean squared error is equal to the sum of squared errors divided by its degrees of freedom. (LOS 10.d)

Module Quiz 10.3

1. **C** The estimate of Y, given $X = 10$, is $Y = 5 + 3.5(10) = 40$. The critical value for a 90% confidence interval with a large sample size (z-statistic) is approximately 1.65. Given the standard error of the forecast of 2.5, the confidence interval for the estimated value of Y is 40 +/− 1.65(2.5) = 35.875 to 44.125. (LOS 10.e)

2. **B** The appropriate model would be a lin-log model, in which the values of the dependent variable (Y) are regressed on the natural logarithms of the independent variable (X): $Y = b_0 + b_1 \ln(X)$. (LOS 10.f)

READING 11

INTRODUCTION TO BIG DATA TECHNIQUES

MODULE 11.1: INTRODUCTION TO FINTECH



Video covering this content is available online.

LOS 11.a: Describe aspects of "fintech" that are directly relevant for the gathering and analyzing of financial data.

The term **fintech** refers to developments in technology that can be applied to the financial services industry. Companies that are in the business of developing technologies for the finance industry are often referred to as fintech companies.

Some of the primary areas where fintech is developing include the following:

- Increasing functionality to handle large sets of data that may come from many sources and exist in various forms
- Tools and techniques such as artificial intelligence for analyzing very large datasets

LOS 11.b: Describe Big Data, artificial intelligence, and machine learning.

Big Data is a widely used expression that refers to all the potentially useful information that is generated in the economy. This includes not only data from traditional sources, such as financial markets, company financial reports, and government economic statistics, but also **alternative data** from nontraditional sources. Some of these nontraditional sources are as follows:

- Individuals generate usable data such as social media posts, online reviews, email, and website visits.
- Businesses generate potentially useful information such as bank records and retail scanner data. These kinds of data are referred to as **corporate exhaust**.
- Sensors, such as radio frequency identification chips, are embedded in numerous devices such as smartphones and smart buildings. The broad network of such devices is referred to as the **Internet of Things**.

Characteristics of Big Data include its volume, velocity, and variety.

The *volume* of data continues to grow by orders of magnitude. The units in which data can be measured have increased from megabytes and gigabytes to terabytes (1,000 gigabytes) and even petabytes (1,000 terabytes).

Velocity refers to how quickly data are communicated. Real-time data such as stock market price feeds are said to have low **latency**. Data that are only communicated periodically or with a lag are said to have high latency.

The *variety* of data refers to the varying degrees of structure in which data may exist. These range from structured forms (e.g., spreadsheets and databases), to semistructured forms (e.g., photos and web page code), to unstructured forms (e.g., video).

The field of **data science** concerns how we extract information from Big Data. Data science describes methods for processing and visualizing data. Processing methods include the following:

- *Capture*. This is collecting data and transforming it into usable forms.
- *Curation*. This is assuring data quality by adjusting for bad or missing data.
- *Storage*. This is archiving and accessing data.
- *Search*. This is examining stored data to find needed information.
- *Transfer*. This is moving data from their source or a storage medium to where they are needed.

Visualization techniques include the familiar charts and graphs that display structured data. To visualize less structured data requires other methods. Some examples of these are word clouds that illustrate the frequency with which words appear in a sample of text, or mind maps that display logical relations among concepts.

Taking advantage of Big Data presents numerous challenges. Analysts must ensure that the data they use are of high quality, accounting for the possibilities of outliers, bad or missing data, or sampling biases. The volume of data collected must be sufficient and appropriate for its intended use.

The need to process and organize data before using it can be especially problematic with qualitative and unstructured data. This is a process to which **artificial intelligence**, or computer systems that can be programmed to simulate human cognition, may be applied usefully. **Neural networks** are an example of artificial intelligence in that they are programmed to process information in a way similar to the human brain.

An important development in the field of artificial intelligence is **machine learning**. In machine learning, a computer algorithm is given inputs of source data, with no assumptions about their probability distributions, and may be given outputs of target data. The algorithm is designed to learn, without human assistance, how to model the output data based on the input data, or to learn how to detect and recognize patterns in the input data.

Machine learning typically requires vast amounts of data. A typical process begins with a *training* dataset in which the algorithm looks for relationships. A *validation* dataset is

then used to refine these relationship models, which can then be applied to a *test* dataset to analyze their predictive ability.

In **supervised learning**, the input and output data are labeled, the machine learns to model the outputs from the inputs, and then the machine is given new data on which to use the model. In **unsupervised learning**, the input data are not labeled, and the machine learns to describe the structure of the data. **Deep learning** is a technique that uses layers of neural networks to identify patterns, beginning with simple patterns and advancing to more complex ones. Deep learning may employ supervised or unsupervised learning. Some of the applications of deep learning include image and speech recognition.

Machine learning can produce models that overfit or underfit the data. **Overfitting** occurs when the machine learns the input and output data too exactly, treats noise as true parameters, and identifies spurious patterns and relationships. In effect, the machine creates a model that is too complex. **Underfitting** occurs when the machine fails to identify actual patterns and relationships, treating true parameters as noise. This means that the model is not complex enough to describe the data. A further challenge with machine learning is that its results can be a "black box," producing outcomes based on relationships that are not readily explainable.

LOS 11.c: Describe applications of Big Data and Data Science to investment management.

Applications of fintech that are relevant to investment management include text analytics, natural language processing, risk governance, and algorithmic trading.

Text analytics refers to the analysis of unstructured data in text or voice forms. An example of text analytics is analyzing the frequency of words and phrases. In the finance industry, text analytics have the potential to partially automate specific tasks such as evaluating company regulatory filings.

Natural language processing refers to the use of computers and artificial intelligence to interpret human language. Speech recognition and language translation are among the uses of natural language processing. Possible applications in finance could be to check for regulatory compliance in an examination of employee communications, or to evaluate large volumes of research reports to detect more subtle changes in sentiment than can be discerned from analysts' recommendations alone.

Risk governance requires an understanding of a firm's exposure to a wide variety of risks. Financial regulators require firms to perform risk assessments and stress testing. The simulations, scenario analysis, and other techniques used for risk analysis require large amounts of quantitative data along with a great deal of qualitative information. Machine learning and other techniques related to Big Data can be useful in modeling and testing risk, particularly if firms use real-time data to monitor risk exposures.

Algorithmic trading refers to computerized securities trading based on a predetermined set of rules. For example, algorithms may be designed to enter the optimal execution instructions for any given trade based on real-time price and volume

data. Algorithmic trading can also be useful for executing large orders by determining the best way to divide the orders across exchanges. Another application of algorithmic trading is **high-frequency trading** that identifies and takes advantage of intraday securities mispricings.



MODULE QUIZ 11.1

1. Fintech is *most accurately* described as the:
   A. application of technology to the financial services industry.
   B. replacement of government-issued money with electronic currencies.
   C. clearing and settling of securities trades through distributed ledger technology.
2. Which of the following technological developments is *most likely* to be useful for analyzing Big Data?
   A. Machine learning.
   B. High-latency capture.
   C. The Internet of Things.

KEY CONCEPTS

LOS 11.a

Fintech refers to developments in technology that can be applied to the financial services industry. Companies that develop technologies for the finance industry are referred to as fintech companies.

LOS 11.b

Big Data refers to the potentially useful information that is generated in the economy, including data from traditional and nontraditional sources. Characteristics of Big Data include its volume, velocity, and variety.

Artificial intelligence refers to computer systems that can be programmed to simulate human cognition. Neural networks are an example of artificial intelligence.

Machine learning is programming that gives a computer system the ability to improve its performance of a task over time and is often used to detect patterns in large sets of data.

LOS 11.c

Applications of fintech to investment management include text analytics, natural language processing, risk governance, and algorithmic trading.

Text analytics refers to analyzing unstructured data in text or voice forms. Natural language processing is the use of computers and artificial intelligence to interpret human language. Algorithmic trading refers to computerized securities trading based on predetermined rules.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 11.1

1. **A** Fintech is the application of technology to the financial services industry and to companies that are involved in developing and applying technology for financial services. Cryptocurrencies and distributed ledger technology are examples of fintech-related developments. (LOS 11.a)

2. **A** Machine learning is a computer programming technique useful for identifying and modeling patterns in large volumes of data. The Internet of Things is the network of devices that is one of the sources of Big Data. Capture is one aspect of processing data. Latency is the lag between when data is generated and when it is needed. (LOS 11.b)

TOPIC QUIZ: QUANTITATIVE METHODS

You have now finished the Quantitative Methods topic section. Please log into your Schweser online dashboard and take the Topic Quiz on this section. The Topic Quiz provides immediate feedback on how effective your study has been for this material. Questions are more exam-like than typical Module Quiz or QBank questions; a score of less than 70% indicates that your study likely needs improvement. These tests are best taken timed; allow 1.5 minutes per question.

READING 12

FIRMS AND MARKET STRUCTURES

MODULE 12.1: BREAKEVEN, SHUTDOWN, AND SCALE

Video covering this content is available online.

LOS 12.a: Determine and interpret breakeven and shutdown points of production, as well as how economies and diseconomies of scale affect costs under perfect and imperfect competition.

In economics, we define the **short run** for a firm as the time period over which some factors of production are fixed. Typically, we assume that capital is fixed in the short run so that a firm cannot change its scale of operations (plant and equipment) over the short run. All factors of production (costs) are variable in the **long run**. The firm can let its leases expire and sell its equipment, thereby avoiding costs that are fixed in the short run.

Shutdown and Breakeven Under Perfect Competition

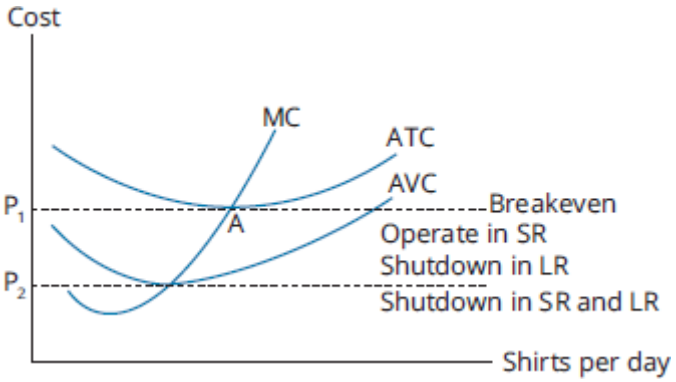
As a simple example of shutdown and breakeven analysis, consider a retail store with a one-year lease (fixed cost) and one employee (quasi-fixed cost), so that variable costs are simply the store's cost of merchandise. If the total sales (total revenue) just cover both fixed and variable costs, price equals both average revenue and average total cost —so we are at the breakeven output quantity, and economic profit equals zero.

During the period of the lease (the short run), as long as items are being sold for more than their variable cost, the store should continue to operate to minimize losses. If items are being sold for less than their average variable cost, losses would be reduced by shutting down the business in the short run.

In the long run, a firm should shut down if the price is less than average total cost, regardless of the relation between price and average variable cost.

For a firm under perfect competition (a price-taker), we can use a graph of cost functions to examine the profitability of the firm at different output prices. In Figure 12.1, at price P_1, price and average revenue equal average total cost. At the output level of Point A, the firm is making an economic profit of zero. At a price above P_1, economic profit is positive, and at prices less than P_1, economic profit is negative (the firm has economic losses).

Figure 12.1: Shutdown and Breakeven



Because some costs are fixed in the short run, it will be better for the firm to continue production in the short run as long as average revenue is greater than average variable costs. At prices between P_1 and P_2 in Figure 12.1, the firm has losses, but the losses are smaller than would occur if all production were stopped. As long as total revenue is greater than total variable cost, at least some of the firm's fixed costs are covered by continuing to produce and sell its product. If the firm were to shut down, losses would be equal to the fixed costs that still must be paid. As long as price is greater than average variable costs, the firm will minimize its losses in the short run by continuing in business.

If average revenue is less than average variable cost, the firm's losses are greater than its fixed costs, and it will minimize its losses by shutting down production in the short run. In this case (a price less than P_2 in Figure 12.1), the loss from continuing to operate is greater than the loss (total fixed costs) if the firm is shut down.

In the long run, all costs are variable, so a firm can avoid its (short-run) fixed costs by shutting down. For this reason, if price is expected to remain below minimum average total cost (Point A in Figure 12.1) in the long run, the firm will shut down rather than continue to generate losses.

To sum up, if average revenue is less than average variable cost in the short run, the firm should shut down. This is its **short-run shutdown point**. If average revenue is greater than average variable cost in the short run, the firm should continue to operate, even if it has losses. In the long run, the firm should shut down if average revenue is less than average total cost. This is the **long-run shutdown point**. If average revenue is just equal to average total cost, total revenue is just equal to total (economic) cost, and this is the firm's **breakeven point**.

- If AR ≥ ATC, the firm should stay in the market in both the short and long run.
- If AR ≥ AVC, but AR < ATC, the firm should stay in the market in the short run but will exit the market in the long run.
- If AR < AVC, the firm should shut down in the short run and exit the market in the long run.

Shutdown and Breakeven Under Imperfect Competition

For price-searcher firms (those that face downward-sloping demand curves), we could compare average revenue to ATC and AVC, just as we did for price-taker firms, to identify shutdown and breakeven points. However, marginal revenue is no longer equal to price.

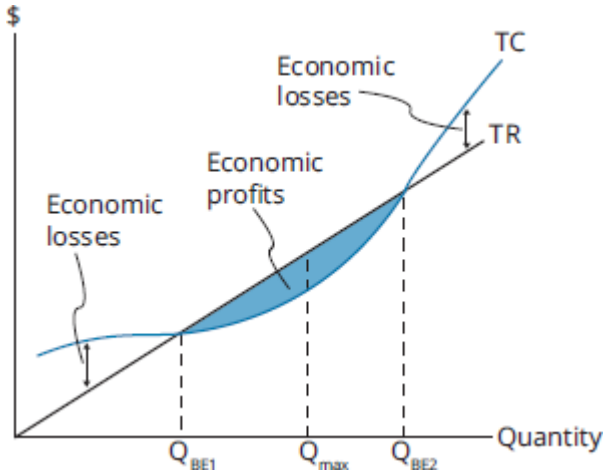
We can, however, still identify the conditions under which a firm is breaking even, should shut down in the short run, and should shut down in the long run in terms of total costs and total revenue. These conditions are as follows:

- TR = TC: break even
- TC > TR > TVC: firm should continue to operate in the short run but shut down in the long run
- TR < TVC: firm should shut down in the short run and the long run

Because price does not equal marginal revenue for a firm in imperfect competition, analysis based on total costs and revenues is better suited for examining breakeven and shutdown points.

The previously described relations hold for both price-taker and price-searcher firms. We illustrate these relations in Figure 12.2 for a price-taker firm (TR increases at a constant rate with quantity). Total cost equals total revenue at the breakeven quantities Q_{BE1} and Q_{BE2}. The quantity for which economic profit is maximized is shown as Q_{max}.

Figure 12.2: Breakeven Point Using the Total Revenue/Total Cost Approach



If the entire TC curve exceeds TR (i.e., no breakeven point), the firm will want to minimize the economic loss in the short run by operating at the quantity corresponding to the smallest (negative) value of TR − TC.

> **EXAMPLE: Short-run shutdown decision**
>
> For the last fiscal year, Legion Gaming reported total revenue of $700,000, total variable costs of $800,000, and total fixed costs of $400,000. Should the firm

continue to operate in the short run?

Answer:

The firm should shut down. Total revenue of $700,000 is less than total costs of $1,200,000, and it is also less than total variable costs of $800,000. By shutting down, the firm will lose an amount equal to fixed costs of $400,000. This is less than the loss of operating, which is TR − TC = $500,000.

EXAMPLE: Long-run shutdown decision

Suppose, instead, that Legion Gaming reported total revenue of $850,000. Should the firm continue to operate in the short run? Should it continue to operate in the long run?

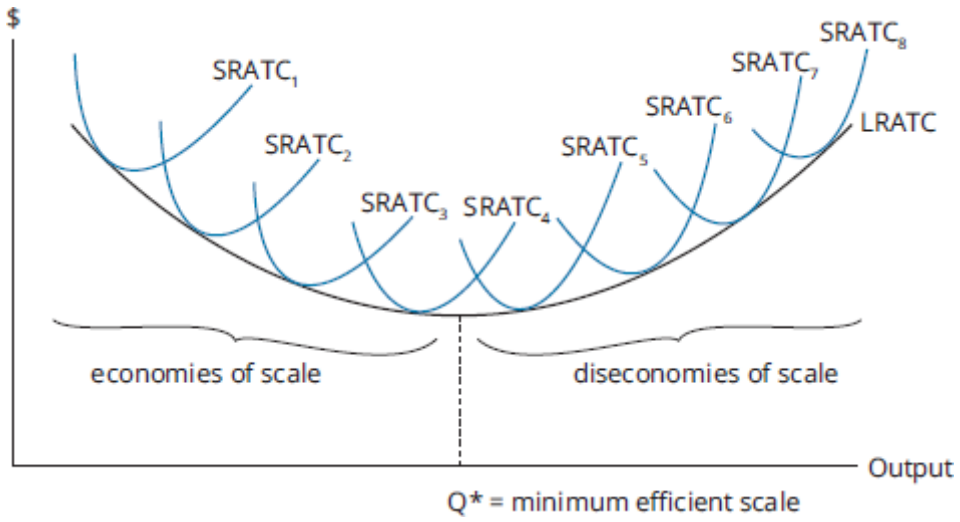
Answer:

In the short run, TR > TVC, and the firm should continue operating. The firm should consider exiting the market in the long run, as TR is not sufficient to cover all of the fixed costs and variable costs.

Economies and Diseconomies of Scale

While plant size is fixed in the short run, in the long run, firms can choose their most profitable scale of operations. Because the long-run average total cost (LRATC) curve is drawn for many different plant sizes or scales of operation, each point along the curve represents the minimum ATC for a given plant size or scale of operations. In Figure 12.3, we show a firm's LRATC curve along with short-run average total cost (SRATC) curves for many different plant sizes, with $SRATC_{n+1}$ representing a larger scale of operations than $SRATC_n$.

Figure 12.3: Economies and Diseconomies of Scale



We draw the LRATC curve as U-shaped. Average total costs first decrease with larger scale, but eventually begin to increase with larger scale. The lowest point on the LRATC corresponds to the scale or plant size at which the average total cost of production is at a minimum. This scale is sometimes called the **minimum efficient scale**. Under perfect competition, firms must operate at minimum efficient scale in long-run equilibrium, and LRATC will equal the market price. Recall that under perfect competition, firms earn zero economic profit in long-run equilibrium. Firms that have chosen a different scale of operations with higher average total costs will have economic losses and must either leave the industry or change to the minimum efficient scale.

The downward-sloping segment of the LRATC curve presented in Figure 12.3 indicates that **economies of scale** (or *increasing returns to scale*) are present. Economies of scale result from factors such as labor specialization, mass production, and investment in more efficient equipment and technology. In addition, the firm may be able to negotiate lower input prices with suppliers as it increases in size and purchases more resources. A firm operating with economies of scale can increase its competitiveness by expanding production and reducing costs.

The upward-sloping segment of the LRATC curve indicates that **diseconomies of scale** are present. Diseconomies of scale may result as the increasing bureaucracy of larger firms leads to inefficiency, problems with motivating a larger workforce, and greater barriers to innovation and entrepreneurial activity. A firm operating under diseconomies of scale will want to decrease output and move back toward the minimum efficient scale. The U.S. auto industry is an example of an industry that has exhibited diseconomies of scale.

There may be a relatively flat portion at the bottom of the LRATC curve that exhibits *constant returns to scale*, or relatively constant costs across a range of plant sizes.



MODULE QUIZ 12.1

1. In a purely competitive market, economic losses indicate that:
   A. price is below average total costs.
   B. collusion is occurring in the marketplace.
   C. firms need to expand output to reduce costs.

2. A firm is likely to operate in the short run as long as price is at least as great as:
   A. marginal cost.
   B. average total cost.
   C. average variable cost.

3. A firm's average revenue is greater than its average variable cost and less than its average total cost. If this situation is expected to persist, the firm should:
   A. shut down in the short run and in the long run.
   B. shut down in the short run, but operate in the long run.
   C. operate in the short run, but shut down in the long run.

4. If a firm increases its plant size by 10% and its minimum average total cost increases by 10%, the firm is experiencing:
   A. constant returns to scale.
   B. diseconomies of scale.

C. economies of scale.

MODULE 12.2: CHARACTERISTICS OF MARKET STRUCTURES

Video covering this content is available online.

LOS 12.b: Describe characteristics of perfect competition, monopolistic competition, oligopoly, and pure monopoly.

Recall from the prerequisite readings that perfect competition results in firm demand that is horizontal (perfectly elastic) at the market price. The firm demand curves for the three other market structures we discuss are all downward sloping. When a firm's demand curve slopes downward, marginal revenue (MR) is less than price. For both horizontal and downward-sloping demand curves, a firm will maximize profits by producing the quantity for which MR is just equal to marginal cost.

While it may not be true in every case, it may be useful to think of firms under pure monopoly as having the steepest demand curves. Firms under monopolistic competition typically have relatively elastic downward-sloping demand curves, while firms in an oligopoly market will face downward-sloping demand curves somewhere between these two extremes.

We can analyze where a market falls along the spectrum from perfect competition to pure monopoly by examining five factors:

1. Number of firms and their relative sizes
2. Degree to which firms differentiate their products
3. Bargaining power of firms with respect to pricing
4. Barriers to entry into or exit from the industry
5. Degree to which firms compete on factors other than price

Perfect competition refers to a market in which many firms produce identical products, barriers to entry into the market are very low, and firms compete for sales only on the basis of price. Firms face perfectly elastic (horizontal) demand curves at the price determined in the market because no firm has a large enough portion of the overall market to affect the market price of the good. The market for wheat in a region is a good approximation of such a market. Overall market supply and demand determine the price of wheat, and each producer can sell all that they choose to at that price.

Monopolistic competition differs from perfect competition in that products are not identical. Each firm differentiates its product(s) from those of other firms through some combination of differences in product quality, product features, and marketing. The demand curve faced by each firm is downward sloping (i.e., neither perfectly elastic nor perfectly inelastic). Prices that producers charge are not identical because of perceived differences among their products, and typically, barriers to entry are low. The market for toothpaste is a good example of monopolistic competition. Firms differentiate their products through features and marketing with claims of more attractiveness, whiter teeth, fresher breath, and even actually cleaning your teeth and preventing decay. If the

price of your personal favorite increases, you are not likely to immediately switch to another brand as we assume under perfect competition. Some customers may switch brands in response to a 10% increase in price, and some may not. This is why firm demand is downward sloping rather than perfectly elastic.

The most important characteristic of an **oligopoly** market is that only a few firms are in the industry. In such a market, each firm must consider the actions and responses of other firms in setting price and business strategy. We say that such firms are *interdependent*. While the firms' products are typically good substitutes for each other, they may be either quite similar or differentiated through features, branding, marketing, and quality. Barriers to entry are typically high, often because of economies of scale in production or marketing, which accounts for the existence of a small number of firms with relatively large market shares. Demand can be more or less elastic than for firms in monopolistic competition. The automobile market is dominated by a small number of large firms and can be characterized as an oligopoly. The product and pricing decisions of Toyota certainly affect those of Ford, and vice versa. Automobile makers compete based on price, but they also compete through marketing, product features, and quality, which are often signaled strongly through the brand name. The oil industry also has a few firms with large market shares, but their products are, in most cases, good substitutes for each other.

A **monopoly** market is characterized by a single seller of a product with no good substitutes. This fact alone means that the firm faces a downward-sloping demand curve (the market demand curve) and has the power to choose the price at which it sells its product. High barriers to entry protect a monopoly producer from competition. One source of monopoly power is the protection offered by copyrights and patents. Another possible source of monopoly power is control over a resource specifically needed to produce the product. Most frequently, monopoly power is supported by specific laws or government regulation (e.g., a local electric utility).

Figure 12.4 shows the key features of each market structure.

Figure 12.4: Characteristics of Market Structures

	Perfect Competition	Monopolistic Competition	Oligopoly	Monopoly
Number of sellers	Many firms	Many firms	Few firms	Single firm
Barriers to entry	Very low	Low	High	Very high
Nature of substitute products	Very good substitutes	Good substitutes, but differentiated	Good substitutes or differentiated	No good substitutes
Nature of competition	Price only	Price, marketing, features	Price, marketing, features	Advertising
Pricing power	None	Some	Some to significant	Significant

LOS 12.c: Explain supply and demand relationships under monopolistic competition, including the optimal price and output for firms as well as pricing strategy.

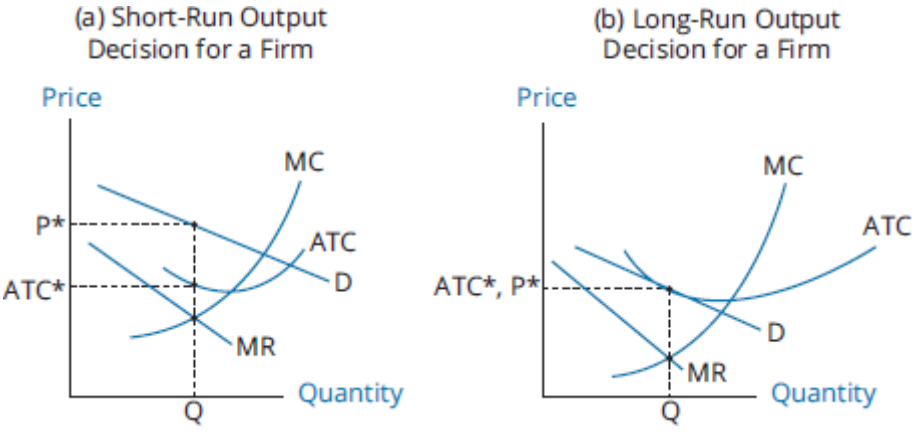
Monopolistic competition has the following market characteristics:

- *A large number of independent sellers*. (1) Each firm has a relatively small market share, so no individual firm has any significant power over price. (2) Firms only need to pay attention to average market price, not the prices of individual competitors. (3) There are too many firms in the industry for collusion (price-fixing) to be possible.
- *Differentiated products*. Each producer has a product that is, in some way, different from those of its competitors (in the minds of consumers). The competing products are considered close substitutes for one another.
- *Firms compete less on price and more on marketing, perceived quality, and differences in features*. Firms must make price and output decisions because they face downward-sloping demand curves.
- *Low barriers to entry*. The cost of entering the market and exiting the market are relatively low.

Think about the market for toothpaste. Brands of toothpaste are quite similar, and it is reasonable to assume that toothpaste is not too difficult or costly to produce. But brands are differentiated based on specific features, on influential advertising and marketing, and on the reputations of the producers.

The price/output decision for monopolistic competition is illustrated in Figure 12.5. Panel A of Figure 12.5 illustrates the short-run price/output characteristics of monopolistic competition for a single firm. As indicated, firms in monopolistic competition maximize economic profits by producing where marginal revenue (MR) equals marginal cost (MC), and by charging the price for that quantity from the demand curve, *D*. Here, the firm earns positive economic profits because price, *P**, exceeds average total cost, *ATC**. Due to low barriers to entry, competitors can enter the market in pursuit of these economic profits.

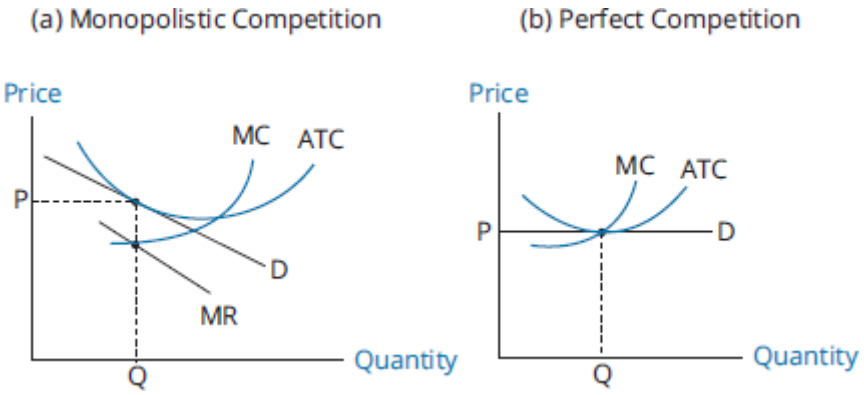
Figure 12.5: Short-Run and Long-Run Output Under Monopolistic Competition



Panel B of Figure 12.5 illustrates long-run equilibrium for a *representative* firm after new firms have entered the market. As indicated, the entry of new firms shifts the demand curve faced by each individual firm down to the point where price equals average total cost (P* = ATC*), such that economic profit is zero. At this point, there is no longer an incentive for new firms to enter the market, and long-run market equilibrium is established. A firm in monopolistic competition continues to produce at the quantity where MR = MC, but it no longer earns positive economic profits. We can get to a similar long-run equilibrium even without entry of new firms if each firm increases its marketing spending (a component of ATC) to either increase or defend its market share until ATC has increased to that shown in Panel B. If all firms compete in this way, each firm will produce Q* and sell at P* but earn no economic profit because marketing costs have increased ATC to ATC*, which is equal to price. Advertising expenses are often relatively high for firms in monopolistic competition.

Figure 12.6 illustrates the differences between long-run equilibrium in markets with monopolistic competition and markets with perfect competition. Note that with monopolistic competition, price is greater than MC (i.e., producers can realize an economic profit), average total cost is not at a minimum for the quantity produced (suggesting excess capacity, or an inefficient scale of production), and the price is slightly higher than under perfect competition. The point to consider here, however, is that perfect competition is characterized by no product differentiation. The question of the efficiency of monopolistic competition becomes, "Is there an economically efficient amount of product differentiation?"

Figure 12.6: Firm Output Under Monopolistic and Perfect Competition



In a world with only one brand of toothpaste, clearly, average production costs would be lower. That fact alone probably does not mean that a world with only one brand or type of toothpaste would be a better world. While product differentiation has costs, it also has benefits to consumers. Consider the market for a pharmaceutical that reduces blood pressure to prolong life. There may be several competing drugs that are more or less effective for, or well or poorly tolerated by, different groups of patients. In this case, we may find that firm demand curves are relatively steep compared to those for brands of toothpaste, indicating that the competing drugs are not considered good substitutes for many patients.

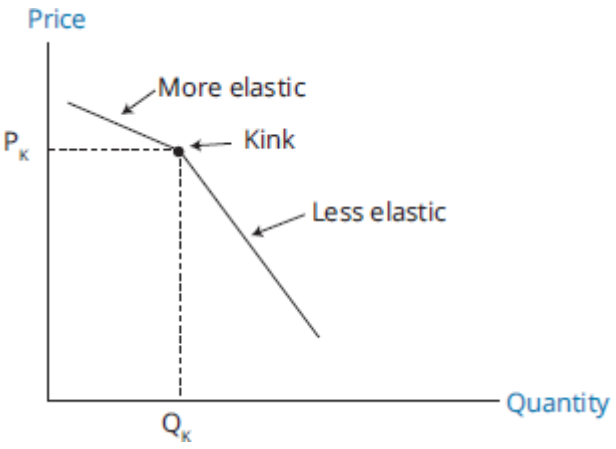
LOS 12.d: Explain supply and demand relationships under oligopoly, including the optimal price and output for firms as well as pricing strategy.

Compared to monopolistic competition, an oligopoly market has higher barriers to entry and fewer firms. The other key difference is that the firms are interdependent; a price change by one firm can be expected to be met by a price change by its competitors in response. This means that the actions of another firm will directly affect a given firm's demand curve for the product. Given this complicating fact, models of oligopoly pricing and profits must make numerous important assumptions. In the following, we describe four of these models and their implications for price and quantity:

1. Kinked demand curve model
2. Cournot duopoly model
3. Nash equilibrium model
4. Stackelberg dominant firm model

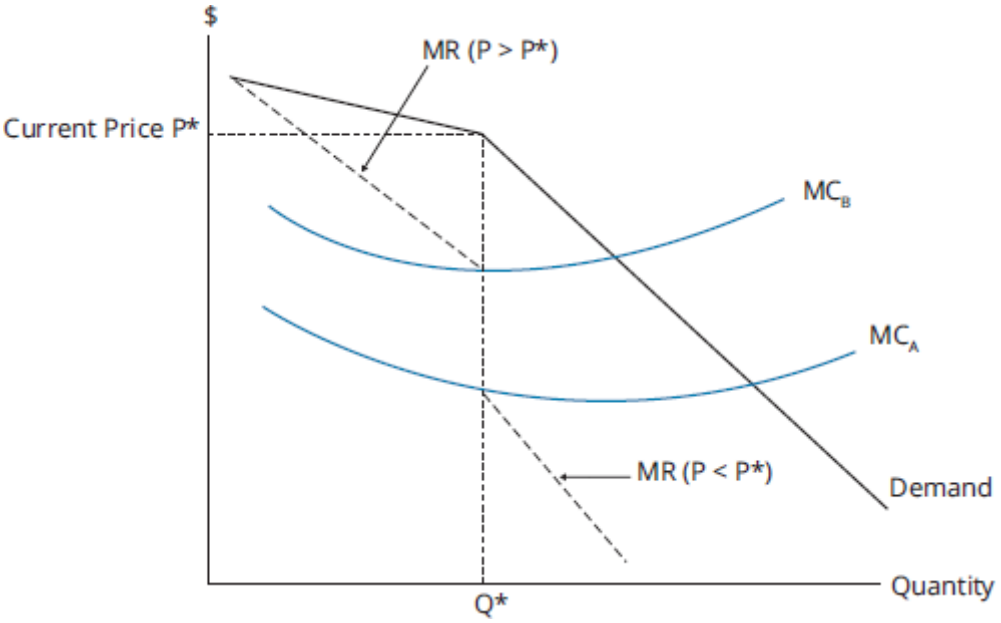
One traditional model of oligopoly, the **kinked demand curve model**, is based on the assumption that competitors are unlikely to match a price increase by a competitor, but very likely to match a price decrease by a competitor. This results in a kink in the demand curves faced by each producer, at the current market price. Each firm believes that it faces a demand curve that is more elastic (flatter) above the current price (the kink in the demand curve) than it is below the given price. The kinked demand curve model is illustrated in Figure 12.7. The kink price is at price P_K, where a firm produces Q_K. A firm believes that if it raises its price above P_K, its competitors will remain at P_K, and it will lose market share because it has the highest price. Above P_K, the demand curve is considered to be relatively elastic (i.e., a small price increase will result in a large decrease in demand). On the other hand, if a firm decreases its price below P_K, other firms will match the price cut, and all firms will experience a relatively small increase in sales relative to any price reduction. Therefore, Q_K is the profit-maximizing level of output.

Figure 12.7: Kinked Demand Curve Model



With a kink in the demand curve, we also get a gap in the associated MR curve, as shown in Figure 12.8. For any firm with a MC curve passing through this gap, the price where the kink is located is the firm's profit-maximizing price.

Figure 12.8: Marginal Revenue With Kinked Demand Curve



We say that the decisions of firms in an oligopoly are interdependent; that is, the pricing decision of one firm depends on the pricing decisions of other firms. Some models of market price equilibrium have a set of rules for the actions of oligopolists. These rules assume they choose prices based on the choices of the other firms. By specifying the decision rules that each firm follows, we can design a model that allows us to determine the equilibrium prices and quantities for firms operating in an oligopoly market.

An early model of oligopoly pricing decisions is the **Cournot model**. In Cournot's duopoly model, two firms with identical MC curves each choose their preferred selling price based on the price the other firm chose in the previous period. Firms assume that the competitor's price will not change. The long-run equilibrium for an oligopoly with two firms (duopoly), in the Cournot model, is for both firms to sell the same quantity, dividing the market equally at the equilibrium price. The equilibrium price is less than the price that a single monopolist would charge, but greater than the equilibrium price that would result under perfect competition. With a greater number of producers, the long-run market equilibrium price moves toward the competitive price.

Another model, the **Stackelberg model**, uses a different set of rules and produces a different result. While the Cournot model assumes the competitors choose price simultaneously each period, the Stackelberg model assumes pricing decisions are made sequentially. One firm, the "leader," chooses its price first, and the other firm chooses a price based on the leader's price. In long-run equilibrium, under these rules, the leader charges a higher price and receives a greater proportion of the firms' total profits.

These models are early versions of *rules-based models*, which fall under the heading of what are now generally termed *strategic games*. Strategic games comprise decision

models in which the best choice for a firm depends on the expected actions (reactions) of other firms.

A more general model of strategic games was developed by Nobel Prize winner John Nash, who developed the concept of a **Nash equilibrium**. A Nash equilibrium is reached when the choices of all firms are such that there is no other choice that makes any firm better off (increases profits or decreases losses). The Cournot model results in a Nash equilibrium. In equilibrium, neither competitor can increase profits by changing the price they charge.

The concept of a Nash equilibrium can be illustrated with the situation presented in Figure 12.9, which shows the choices and resulting profits for two firms. Each firm can charge either a high price or a low price. If both firms charge a high price, Firm A earns 1,000 and Firm B earns 600. While Firm A would not charge the low price (it would earn less regardless of Firm B's decision), Firm B can increase profits to 700 by charging a low price. With Firm A charging a high price and Firm B charging a low price, neither firm can increase profits by unilaterally changing its price strategy. Thus, we can identify the Nash equilibrium in this scenario as Firm B charging a low price and Firm A charging a high price.

Figure 12.9: Nash Equilibrium

	Firm B high price	Firm B low price
Firm A high price	A earns 1,000 B earns 600	A earns 600 B earns 700
Firm A low price	A earns 160 B earns 0	A earns 100 B earns 140

Thus far, we have assumed that firms act competitively to maximize their individual profits. We can illustrate the potential of cooperative behavior among producers to increase the total profits of all firms in the market. **Collusion** refers to competitors making a joint agreement to charge a given price—or, alternatively, to agree to specific levels of output. In Figure 12.9, the greatest joint profits (1,600) are earned when both firms charge a high price. If Firm A offers to pay Firm B 200 for charging a high price, Firm A's profits increase from 600 to 1,000. After paying 200 to Firm B, Firm A still gains 200. Firm B's profits (including the payment of 200) increase from 700 to 800. Collusion, in this case, increases the profits of both firms, compared to the Nash equilibrium. If firms can enter into and enforce an agreement regarding pricing and output, often they can all benefit. Such agreements among producers are illegal in many countries because they reduce competition.

An example of a collusive agreement is the OPEC **cartel**. Cartel-member countries agree to restrict their oil production to increase the world price of oil. Members sometimes choose to "cheat" on the cartel agreement by producing more than the amount of oil they have agreed to produce. If members of a cartel do not adhere to the agreement (taking advantage of the higher market price but failing to restrict output to the agreed-upon amount), the agreement can quickly break down.

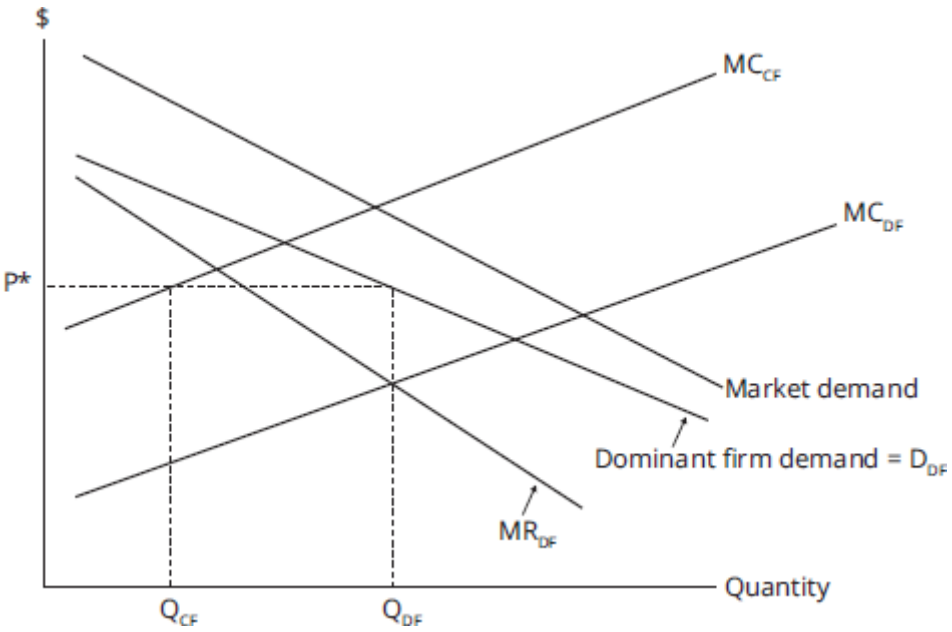
In general, collusive agreements to increase price in an oligopoly market will be more successful (have less cheating) under the following conditions:

- There are fewer firms.
- Products are more similar (less differentiated).
- Cost structures are more similar.
- Purchases are relatively small and frequent.
- Retaliation by other firms for cheating is more certain and more severe.
- There is less actual or potential competition from firms outside the cartel.

A final model of oligopoly behavior to consider is the **dominant firm model**. In this model, a single firm has a significantly large market share because of its greater scale and lower cost structure—the dominant firm (DF). In such a model, the market price is essentially determined by the DF, and the other competitive firms (CFs) take this market price as given.

The DF believes that the quantity supplied by the other firms decreases at lower prices, so that the DF's demand curve is related to the market demand curve, as shown in Figure 12.10. Based on this demand curve (D_{DF}) and its associated marginal revenue (MR_{DF}) curve, the firm will maximize profits at a price of P^*. The CFs maximize profits by producing the quantity for which their marginal cost (MC_{CF}) equals P^*, quantity Q_{CF}.

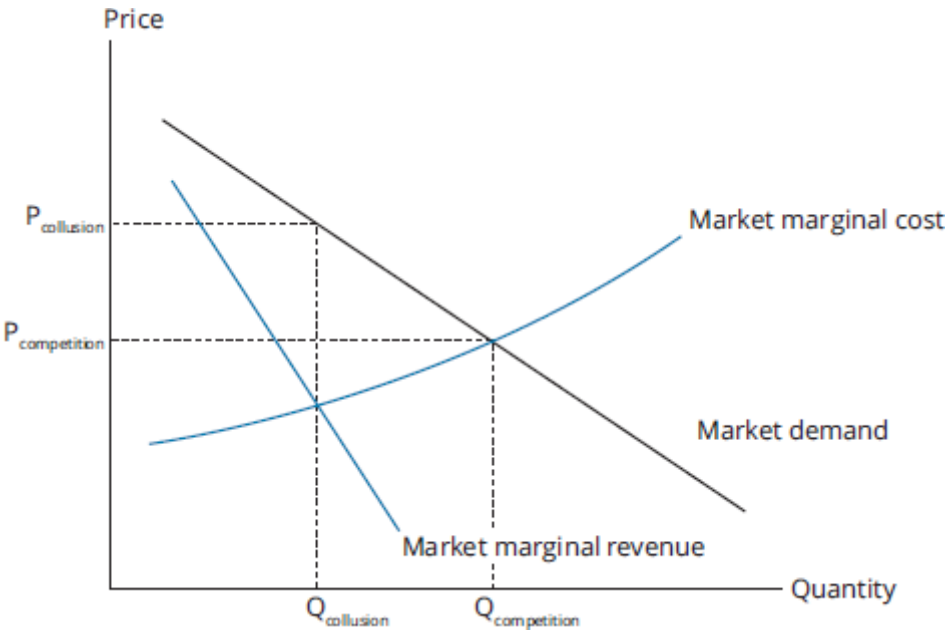
Figure 12.10: Dominant Firm Oligopoly



A price decrease by one of the CFs, which increases Q_{CF} in the short run, will lead to a decrease in price by the DF, and CFs will decrease output or exit the industry in the long run. The long-run result of such a price decrease by competitors below P^* would then be to decrease the overall market share of competitor firms and increase the market share of the DF.

Clearly, oligopoly markets exhibit many possible outcomes that depend on the characteristics of the firms and of the market itself. The important point is that the firms' decisions are interdependent so that the expected reactions of other firms are an important consideration. Overall, the resulting price will be somewhere between the price based on perfect collusion that would maximize total profits to all firms in the market (which is actually the monopoly price), and the price that would result from perfect competition and generate zero economic profits in the long run. These two limiting outcomes are illustrated in Figure 12.11 as $P_{\text{collusion}}$ with $Q_{\text{collusion}}$ for perfect collusion, and $P_{\text{competition}}$ and $Q_{\text{competition}}$ for perfect competition.

Figure 12.11: Collusion vs. Perfect Competition



MODULE QUIZ 12.2

1. The demand for products from monopolistic competitors is relatively elastic due to:
   A. high barriers to entry.
   B. the availability of many close substitutes.
   C. the availability of many complementary goods.

2. Compared to a perfectly competitive industry, in an industry characterized by monopolistic competition:
   A. both price and quantity are likely to be lower.
   B. price is likely to be higher, and quantity is likely to be lower.
   C. quantity is likely to be higher, and price is likely to be lower.

3. A firm will *most likely* maximize profits at the quantity of output for which:
   A. price equals marginal cost.
   B. price equals marginal revenue.
   C. marginal cost equals marginal revenue.

4. An oligopolistic industry has:
   A. few barriers to entry.
   B. few economies of scale.
   C. a great deal of interdependence among firms.

5. Consider a firm in an oligopoly market that believes the demand curve for its product is more elastic above a certain price than below this price. This belief fits *most*

appropriately to which of the following models?

A. Cournot model.
B. Dominant firm model.
C. Kinked demand model.

6. Consider an agreement between France and Germany that will restrict wine production so that maximum economic profit can be realized. The possible outcomes of the agreement are presented in the following table.

	Germany complies	Germany defaults
France complies	France gets €8 billion Germany gets €8 billion	France gets €2 billion Germany gets €10 billion
France defaults	France gets €10 billion Germany gets €2 billion	France gets €4 billion Germany gets €4 billion

Based on the concept of a Nash equilibrium, the *most likely* strategy followed by the two countries with respect to whether they comply with or default on the agreement will be:

A. both countries will default.
B. both countries will comply.
C. one country will default and the other will comply.

MODULE 12.3: IDENTIFYING MARKET STRUCTURES



Video covering this content is available online.

LOS 12.e: Identify the type of market structure within which a firm operates and describe the use and limitations of concentration measures.

We can use the characteristics we outlined earlier to identify the type of market structure within which a firm is operating. Our earlier table is repeated here in Figure 12.12. For an analyst attempting to determine the degree of pricing power that firms in the industry have, the focus is on the number of firms in the industry, its barriers to entry, the nature of substitute products, and the nature of industry competition. Significant interdependence among firm pricing and output decisions is a characteristic of all oligopoly markets, although some interdependence may be present under monopolistic competition—even with many more firms than for an oligopoly structure.

The following table illustrates the differences in characteristics among the various market structures.

Figure 12.12: Characteristics of Market Structures

	Perfect Competition	Monopolistic Competition	Oligopoly	Monopoly
Number of sellers	Many firms	Many firms	Few firms	Single firm
Barriers to entry	Very low	Low	High	Very high
Nature of substitute products	Very good substitutes	Good substitutes, but differentiated	Good substitutes or differentiated	No good substitutes
Nature of competition	Price only	Price, marketing, features	Price, marketing, features	Advertising
Pricing power	None	Some	Some to significant	Significant

Given the differences in cost structures, suitability of substitutes, and the degree of differentiation, simply classifying the primary characteristics of a market does not tell us the degree of pricing power for individual firms or the magnitude of the difference between market prices and the prices implied by perfect competition. What we would really like (especially for the regulation of markets) is to be able to measure the elasticity of firm demand directly, but that is difficult and subject to estimation error.

Consequently, regulators often use market shares (percentages of market sales) to measure the degree of monopoly or market power of firms in an industry. Often, mergers or acquisitions of companies in the same industry or market are not permitted by government authorities when they determine that the market share of the combined firms will be too high and, therefore, detrimental to the economy.

Market or industry **concentration measures** are often used as an indicator of market power. One concentration measure is the **N-firm concentration ratio**, which is calculated as the sum of the percentage market shares of the largest *N* firms in a market. While this measure is simple to calculate and understand, it does not directly measure market power or elasticity of demand.

One limitation of the N-firm concentration ratio is that it may be relatively insensitive to mergers of firms within an industry. This problem is reduced by using an alternative measure of market concentration, the **Herfindahl-Hirschman Index (HHI)**. The HHI is calculated as the sum of the squares of the market shares of the largest firms in the market. The following example illustrates this difference between the two measures and their calculation.

EXAMPLE: 4-firm concentration ratio and 4-firm HHI

Given the market shares of the following firms, calculate the 4-firm concentration ratio and the 4-firm HHI, both before and after a merger of Acme and Blake.

Firm	Market Share
Acme	25%
Blake	15%
Curtis	15%
Dent	10%
Erie	5%
Federal	5%

Answer:

Before the merger, the 4-firm concentration ratio for the market is 25 + 15 + 15 + 10 = 65%. After the merger, the Acme + Blake firm has 40% of the market, and the 4-firm concentration ratio is 40 + 15 + 10 + 5 = 70%. Although the 4-firm concentration ratio has only increased slightly, the market power of the largest firm in the industry has increased significantly, from 25% to 40%.

Before the merger, the 4-firm HHI is $0.25^2 + 0.15^2 + 0.15^2 + 0.10^2 = 0.1175$.

After the merger, the 4-firm HHI is $0.40^2 + 0.15^2 + 0.10^2 + 0.05^2 = 0.1950$, a significant increase.

A limitation that applies to both of our simple concentration measures is that barriers to entry are not considered in either case. Even a firm with high market share may not have much pricing power if barriers to entry are low and there is *potential competition*. With low barriers to entry, it may be the case that other firms stand ready to enter the market if firms currently in the market attempt to increase prices significantly. In this case, the elasticity of demand for existing firms may be high, even though they have relatively high market shares and industry concentration measures.



MODULE QUIZ 12.3

1. Which of the following is *most likely* an advantage of the Herfindahl-Hirschman Index (HHI) relative to the *N*-firm concentration ratio?
   A. The HHI is simpler to calculate.
   B. The HHI considers barriers to entry.
   C. The HHI is more sensitive to mergers.
2. A market characterized by low barriers to entry, good substitutes, limited pricing power, and marketing of product features is *best* characterized as:
   A. oligopoly.
   B. perfect competition.
   C. monopolistic competition.

KEY CONCEPTS

LOS 12.a

The breakeven quantity of production is the quantity for which price (P) = average total cost (ATC), and total revenue (TR) = total cost (TC).

A firm should shut down in the long run if P < ATC so that TR < TC. A firm should shut down in the short run (and the long run) if P < average variable cost (AVC) so that TR < total variable cost (TVC).

The long-run average total cost (LRATC) curve shows the minimum average total cost for each level of output, assuming that the plant size (scale of the firm) can be adjusted. A downward-sloping segment of an LRATC curve indicates economies of scale (increasing returns to scale). Over such a segment, increasing the scale of the firm reduces ATC. An upward-sloping segment of an LRATC curve indicates diseconomies of scale, where average unit costs will rise as the scale of the business (and long-run output) increases.

LOS 12.b

Perfect competition is characterized by the following:

- Many firms, each small relative to the market
- Very low barriers to entry into or exit from the industry
- Homogeneous products that are perfect substitutes; no advertising or branding
- No pricing power

Monopoly is characterized by the following:

- A single firm that comprises the whole market
- Very high barriers to entry into or exit from the industry
- Advertising used to compete with substitute products
- Significant pricing power

Monopolistic competition is characterized by the following:

- Many firms
- Low barriers to entry into or exit from the industry
- Differentiated products; heavy advertising and marketing expenditure
- Some pricing power

Oligopoly markets are characterized by the following:

- Few sellers
- High barriers to entry into or exit from the industry
- Products that may be homogeneous or differentiated by branding and advertising
- Firms that may have significant pricing power

LOS 12.c

Under monopolistic competition, firms face downward-sloping demand curves so that marginal revenue is less than price, and the price from the demand curve at the profit-maximizing quantity is the price at the optimal (profit-maximizing) level of output. Resources expended on product differentiation may increase ATC so that there are no economic profits at long-run equilibrium. There is no well-defined firm supply curve.

Monopolistic competition:

- Price > marginal revenue = marginal cost (in equilibrium)
- Zero economic profit in long-run equilibrium

LOS 12.d

Under oligopoly, the pricing strategy is not clear. Because firm decisions are interdependent, the optimal pricing and output strategy depends on assumptions made about other firms' cost structures and about competitors' responses to a firm's price changes.

Models of oligopoly pricing include the following:

- *Cournot*. Firms with a given MC that make pricing decisions simultaneously share the market equally in LR equilibrium.
- *Stackelberg*. When firm pricing decisions are sequential, the first firm to set its price (price leader) will maintain a larger share of the market, even in the long run.
- *Dominant firm*. A firm with a significantly lower cost of production will gain a disproportionate share of the market and essentially set the price that other competitors can charge.

In general, in the absence of collusion, the long-run oligopoly equilibrium is Nash equilibrium, one in which no competitor can unilaterally change price to increase its profits. There is no well-defined firm supply curve.

Oligopoly:

- Price > marginal revenue = marginal cost (in equilibrium)
- May have positive economic profit in long-run equilibrium, but may move toward zero economic profit over time

LOS 12.e

To identify the market structure in which a firm is operating, we can examine the number of firms in its industry, the prevalence of nonprice competition, and barriers to entry, then compare these to the characteristics that define each market structure. However, none of this provides us with the elasticity of firm demand.

Because direct estimation of the elasticity of firm demand is statistically difficult and imprecise, industry concentration measures are often used as a proxy for the degree of competition. A concentration ratio for N firms is calculated as the percentage of market sales accounted for by the N largest firms in the industry and is used as a simple measure of market structure and market power.

The Herfindahl-Hirschman Index measure of concentration is calculated as the sum of the squared market shares of the largest N firms in an industry and better reflects the effect of mergers on industry concentration.

Neither measure indicates market power directly. Both can be misleading when potential competition restricts pricing power.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 12.1

1. **A** In a purely competitive market, economic losses indicate that firms are overproducing, causing prices to fall below average total costs. This can occur in the short run. In the long run, however, market supply will decrease as firms exit the industry, and prices will rise to the point where economic profits are zero. (LOS 12.a)
2. **C** If price is greater than average variable cost, a firm will continue to operate in the short run because it is covering at least some of its fixed costs. (LOS 12.a)
3. **C** If a firm is generating sufficient revenue to cover its variable costs and part of its fixed costs, it should continue to operate in the short run. If average revenue is likely to remain below average total costs in the long run, the firm should shut down. (LOS 12.a)
4. **B** If minimum average total costs increase as plant size is increased, the firm is experiencing diseconomies of scale. (LOS 12.a)

Module Quiz 12.2

1. **B** The demand for products from firms competing in monopolistic competition is relatively elastic due to the availability of many close substitutes. If a firm increases its product price, it will lose customers to firms selling substitute products at lower prices. (LOS 12.c)
2. **B** Monopolistic competition is likely to result in a higher price and lower quantity of output compared to perfect competition. (LOS 12.c)
3. **C** The profit-maximizing output is the quantity at which marginal revenue equals marginal cost. In a price-searcher industry structure (i.e., any structure that is not perfect competition), price is greater than marginal revenue. (LOS 12.c)
4. **C** An oligopolistic industry has a great deal of interdependence among firms. One firm's pricing decisions or advertising activities will affect the other firms. (LOS 12.d)
5. **C** The kinked demand model assumes that each firm in a market believes that at some price, demand is more elastic for a price increase than for a price decrease. (LOS 12.d)
6. **A** The Nash equilibrium results when each nation pursues the strategy that is best, given the strategy that is pursued by the other nation.
   - Given that Germany complies with the agreement: France will get €8 billion if it complies, but €10 billion if it defaults. Therefore, France should default.

- Given that Germany defaults: France will get €2 billion if it complies, but €4 billion if it defaults. Therefore, France should default.

Because France is better off in either case by defaulting, France will default. Germany will follow the same logic and reach the same conclusion. (LOS 12.d)

Module Quiz 12.3

1. **C** Although the *N*-firm concentration ratio is simple to calculate, it can be relatively insensitive to mergers between companies with large market shares. Neither the HHI nor the *N*-firm concentration ratio consider barriers to entry. (LOS 12.e)

2. **C** These characteristics are associated with a market structure of monopolistic competition. Firms in perfect competition do not compete on product features. Oligopolistic markets have high barriers to entry. (LOS 12.e)

READING 13

UNDERSTANDING BUSINESS CYCLES

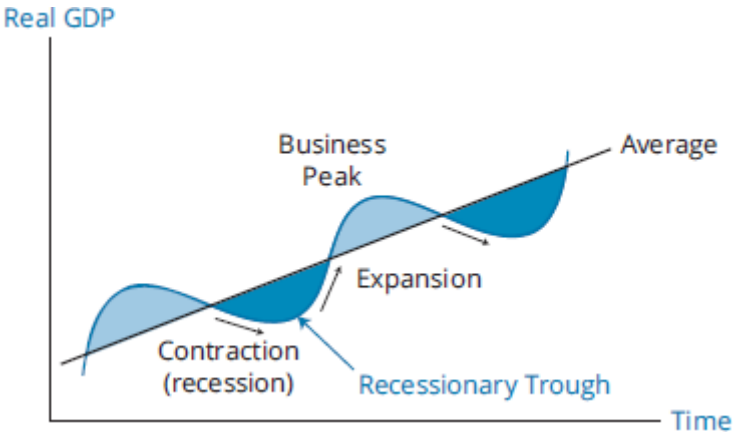
MODULE 13.1: BUSINESS CYCLES

Video covering this content is available online.

LOS 13.a: Describe the business cycle and its phases.

The **business cycle** is characterized by fluctuations in economic activity. The business cycle has four phases: **expansion** (real GDP is increasing), **peak** (real GDP stops increasing and begins decreasing), **contraction** or **recession** (real GDP is decreasing), and **trough** (real GDP stops decreasing and begins increasing). The phases are illustrated in Figure 13.1.

Figure 13.1: Business Cycle



There are alternative ways to show business cycles. The curve in the figure illustrates the **classical cycle**, which is based on real GDP relative to a beginning value. The **growth cycle** refers to changes in the percentage difference between real GDP and its longer-term trend or potential value (shown as the "average" line in the figure). The **growth rate cycle** refers to changes in the annualized percentage growth rate from one period to the next and tends to show both peaks and troughs earlier than the other two measures. The growth rate cycle—which, like the growth cycle, shows GDP relative to a trend rate—is the preferred measure for economists and practitioners.

An expansion features growth in most sectors of the economy, with increasing employment, consumer spending, and business investment. As an expansion approaches

its peak, the rates of increase in spending, investment, and employment slow but remain positive, while inflation accelerates.

A contraction or recession is associated with declines in most sectors, with inflation typically decreasing. When a contraction reaches a trough and the economy begins a new expansion or **recovery**, economic growth becomes positive again and inflation is typically moderate, but employment growth might not start to increase until the expansion has taken hold convincingly.

Economists commonly consider two consecutive quarters of growth in real GDP to be the beginning of an expansion, and two consecutive quarters of declining real GDP to be the beginning of a contraction. Statistical agencies that date expansions and recessions, such as the National Bureau of Economic Research in the United States, look at a wider variety of additional economic data—especially unemployment, industrial production, and inflation—to identify turning points in the business cycle.

A key aspect of business cycles is that they recur, but not at regular intervals. Past business cycles have been as short as a year, or longer than a decade. The idea of a business cycle applies to economies that consist mainly of businesses. For economies that are mostly subsistence agriculture or dominated by state planning, fluctuations in activity are not really "business cycles" in the sense we are discussing here.

LOS 13.b: Describe credit cycles.

Credit cycles refer to cyclical fluctuations in interest rates and the availability of loans (credit). Typically, lenders are more willing to lend, and tend to offer lower interest rates, during economic expansions. Conversely, they are less willing to lend, and require higher interest rates, when the economy is slowing (contracting).

Credit cycles may amplify business cycles. Widely available or "loose" credit conditions during expansions can lead to "bubbles" (prices based on implausible expectations) in the markets for some assets, such as subprime mortgages in the period leading up to the financial crisis of 2007–2009. Some research suggests that expansions tend to be stronger, and contractions deeper and longer lasting, when they coincide with credit cycles. They do not always coincide, however, as historical data suggest credit cycles have been longer in duration than business cycles on average.

LOS 13.c: Describe how resource use, consumer and business activity, housing sector activity, and external trade sector activity vary over the business cycle and describe their measurement using economic indicators.

Business Cycles and Resource Use Fluctuation

Inventories are an important business cycle indicator. Firms try to keep enough inventory on hand to meet sales demand but do not want to keep too much of their capital tied up in inventory. As a result, the **inventory-sales ratio** in many industries tends toward a normal level in times of steady economic growth.

When an expansion is approaching its peak, sales growth begins to slow, and unsold inventories accumulate. This can be seen in an increase in the inventory-sales ratio above its normal level. Firms respond to this unplanned increase in inventory by reducing production, which is one of the causes of the subsequent contraction in the economy. An increase in inventories is counted in the GDP statistics as economic output whether the increase is planned or unplanned. An analyst who looks only at GDP growth, rather than the inventory-sales ratio, might see economic strength rather than the beginning of weakness.

The opposite occurs when a contraction reaches its trough. Having reduced their production levels to adjust for lower sales demand, firms find their inventories becoming depleted more quickly once sales growth begins to accelerate. This causes the inventory-sales ratio to decrease below its normal level. To meet the increase in demand, firms will increase output, and the inventory-sales ratio will increase back toward normal levels.

One of the ways that firms react to fluctuations in business activity is by adjusting their use of labor and physical capital. Adding and subtracting workers in lockstep with changes in economic growth would be costly for firms, in terms of both direct expenses and the damage it would do to employee morale and loyalty. Instead, firms typically begin by changing how they use their current workers, producing less or more output per hour or adjusting the hours they work by adding or removing overtime. Only when an expansion or contraction appears likely to persist will they hire or lay off workers.

Similarly, because it is costly to adjust production levels by frequently buying and selling plant and equipment, firms first adjust their production levels by using their existing physical capital more or less intensively. As an expansion persists, firms will increase their production capacity by investing more in plant and equipment. During contractions, however, firms will not necessarily sell plant and equipment outright. They can reduce their physical capacity by spending less on maintenance or by delaying the replacement of equipment that is nearing the end of its useful life.

Consumer Sector Activity

Consumer spending, the largest component of gross domestic product, depends on the level of consumers' current incomes and their expectations about their future incomes. As a result, consumer spending increases during expansions and decreases during contractions.

Consumer spending in some sectors is more sensitive to business cycle phases than spending in other sectors. Spending on **durable goods** is highly cyclical because they are often higher-value purchases. Consumers are more willing to purchase high-value durable goods (e.g., appliances, furniture, automobiles) during expansions, when incomes are increasing and economic confidence is high. During contractions (and sometimes extending into the early stages of expansions), consumers often postpone durable goods purchases until they are more confident about their employment status and prospects for income growth.

Consumer spending on **services** is also positively correlated with business cycle phases, but not to the same extent as durable goods spending. Services include some

discretionary spending, such as for travel, lodging, and restaurant meals, but they also include spending that is less discretionary, such as for telecommunications, health care, and insurance. Spending on **nondurable goods**, such as food at home or household products for everyday use, remains relatively stable over the business cycle.

Housing Sector Activity

Although the housing sector is a smaller part of the economy relative to overall consumer spending, cyclical swings in activity in the housing market can be large, so that the effect on overall economic activity is greater than it otherwise would be. Important determinants of the level of economic activity in the housing sector are as follows:

1. *Mortgage rates.* Low interest rates tend to increase home buying and construction, while high interest rates tend to reduce home buying and construction.
2. *Housing costs relative to income.* When incomes are cyclically high (low) relative to home costs, including mortgage financing costs, home buying and construction tend to increase (decrease). Housing activity can decrease even when incomes are rising if home prices are rising faster than incomes, as often occurs late in expansions.
3. *Speculative activity.* As we saw in the housing sector in 2007–08 in many economies, rising home prices can lead to purchases based on expectations of further gains. Higher prices led to more construction and eventually excess building. This resulted in falling prices that decreased or eliminated speculative demand and led to dramatic decreases in house prices and in housing activity overall.
4. *Demographic factors.* The proportion of the population in the 25- to 40-year-old segment is positively related to activity in the housing sector because these are the ages of greatest household formation. Strong population shifts from rural areas to cities, as may occur in newly industrializing economies, may require large increases in construction of new housing to accommodate those needs.

External Trade Sector Activity

The most important factors determining the level of a country's imports and exports are domestic GDP growth, GDP growth of trading partners, and currency exchange rates. Increasing growth of domestic GDP leads to increases in purchases of foreign goods (imports), while decreasing domestic GDP growth reduces imports. Exports depend on the growth rates of GDP of other economies (especially those of important trading partners). Increasing foreign incomes increase sales to foreigners (exports), and decreasing economic growth in foreign countries decreases domestic exports.

An increase in the value of a country's currency makes its goods more expensive to foreign buyers and foreign goods less expensive to domestic buyers, which tends to decrease exports and increase imports. A decrease in the value of a country's currency has the opposite effect (increasing exports and decreasing imports). Currencies affect import and export volumes over time in response to persistent trends in foreign exchange rates, rather than in response to short-term changes, which can be quite volatile.

Currency effects on imports and exports can differ in direction from GDP growth effects and respond to a complex set of variables. The effects of changes in GDP levels and growth rates are more direct and immediate.

Typical business cycle characteristics may be summarized as follows:

- *Trough:*
  - The GDP growth rate changes from negative to positive.
  - There is a high unemployment rate, and an increasing use of overtime and temporary workers.
  - Spending on consumer durable goods and housing may increase.
  - There is a moderate or decreasing inflation rate.
- *Expansion:*
  - The GDP growth rate increases.
  - The unemployment rate decreases as hiring accelerates.
  - There are investment increases in producers' equipment and home construction.
  - The inflation rate may increase.
  - Imports increase as domestic income growth accelerates.
- *Peak:*
  - The GDP growth rate decreases.
  - The unemployment rate decreases, but hiring slows.
  - Consumer spending and business investments grow at slower rates.
  - The inflation rate increases.
- *Contraction/recession:*
  - The GDP growth rate is negative.
  - Hours worked decrease; unemployment rate increases.
  - Consumer spending, home construction, and business investments decrease.
  - The inflation rate decreases with a lag.
  - Imports decrease as domestic income growth slows.

Economic Indicators

Economic indicators can be classified into three categories: **leading indicators** that have been known to change direction before peaks or troughs in the business cycle, **coincident indicators** that change direction at roughly the same time as peaks or troughs, and **lagging indicators** that tend not to change direction until after expansions or contractions are already underway.



MODULE QUIZ 13.1

1. In the early part of an economic expansion, inventory-sales ratios are *most likely* to:
   A. increase because sales are unexpectedly low.
   B. increase because businesses plan for expansion.
   C. decrease because of unexpected increases in sales.
2. The contraction phase of the business cycle is *least likely* accompanied by decreasing:
   A. unemployment.
   B. inflation pressure.

C. economic output.

3. Which economic sector would *most likely* correlate strongly and positively with credit cycles?
   A. Exports.
   B. Food retail.
   C. Construction.
4. An economic indicator that has turning points that tend to occur after the turning points in the business cycle is classified as a:
   A. lagging indicator.
   B. leading indicator.
   C. trailing indicator.
5. When they recognize the beginning of a recession, companies are *most likely* to adjust their stock of physical capital by:
   A. selling physical assets.
   B. deferring maintenance of equipment.
   C. canceling orders for new construction equipment.

KEY CONCEPTS

LOS 13.a

The business cycle has four phases:

1. *Expansion*. Real GDP is increasing (or GDP growth relative to trend is increasing).
2. *Peak*. Real GDP stops increasing/begins decreasing (GDP growth relative to trend peaks).
3. *Contraction*. Real GDP is decreasing (or GDP growth relative to trend is decreasing).
4. *Trough*. Real GDP stops decreasing/begins increasing (GDP growth relative to trend reaches a low).

Expansions feature increasing output, employment, consumption, investment, and inflation. Contractions are characterized by decreases in these indicators.

Business cycles are recurring, but they do not occur at regular intervals, can differ in strength or severity, and do not persist for specific lengths of time.

LOS 13.b

Credit cycles are cyclical fluctuations in interest rates and credit availability. Credit cycles may amplify business cycles and cause bubbles in the markets for some assets.

LOS 13.c

Inventory to sales ratios typically increase late in expansions when sales slow, and decrease near the end of contractions when sales begin to accelerate. Firms decrease or increase production to restore their inventory-sales ratios to their desired levels.

Because hiring and laying off employees have high costs, firms prefer to adjust their use of current employees. As a result, firms are slow to lay off employees early in contractions and slow to add employees early in expansions.

Firms use their physical capital more intensively during expansions, investing in new capacity only if they believe the expansion is likely to continue. They use physical capital less intensively during contractions, but they are more likely to reduce capacity by deferring maintenance and not replacing equipment than by selling their physical capital.

Consumer spending fluctuates with the business cycle. Durable goods spending is highly sensitive to business cycles, and spending on services is somewhat sensitive, but spending on nondurable goods is relatively less sensitive to business cycles.

The level of activity in the housing sector is affected by mortgage rates, demographic changes, the ratio of income to housing prices, and investment or speculative demand for homes resulting from recent price trends.

Domestic imports tend to rise with increases in GDP growth and domestic currency appreciation, while increases in foreign incomes and domestic currency depreciation tend to increase domestic export volumes.

Leading indicators have turning points that tend to precede those of the business cycle.

Coincident indicators have turning points that tend to coincide with those of the business cycle.

Lagging indicators have turning points that tend to occur after those of the business cycle.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 13.1

1. **C** Early in an expansion, inventory-sales ratios typically decrease below their normal levels as accelerating sales draw down inventories of produced goods. (LOS 13.c)

2. **A** An economic contraction is likely to feature increasing unemployment (i.e., decreasing employment), along with declining economic output and decreasing inflation pressure. (LOS 13.a)

3. **C** Credit cycles are associated with interest rates and the availability of credit, which is important in the financing of construction and the purchase of property. (LOS 13.b)

4. **A** Lagging indicators have turning points that occur after business cycle turning points. (LOS 13.c)

5. **B** Physical capital adjustments to downturns typically are made through aging of equipment plus postponing maintenance. (LOS 13.c)

READING 14

FISCAL POLICY

MODULE 14.1: FISCAL POLICY OBJECTIVES



Video covering this content is available online.

LOS 14.a: Compare monetary and fiscal policy.

Fiscal policy refers to a government's use of spending and taxation to influence economic activity. The budget is said to be *balanced* when tax revenues equal government expenditures. A **budget surplus** occurs when government tax revenues exceed expenditures, and a **budget deficit** occurs when government expenditures exceed tax revenues. An increase in the deficit (or a decrease in a surplus) is considered expansionary in that it tends to increase GDP. A decrease in a deficit (or increase in a surplus) is considered contractionary in that it tends to decrease GDP.

Monetary policy refers to the central bank's actions that affect the quantity of money and credit in an economy to influence economic activity. Monetary policy is said to be expansionary (or accommodative or easy) when the central bank increases the quantity of money and credit in an economy. Conversely, when the central bank is reducing the quantity of money and credit in an economy, monetary policy is said to be contractionary (or restrictive or tight).

Policymakers use both monetary and fiscal policies with the goals of maintaining stable prices and producing positive economic growth. Fiscal policy can also be used as a tool for redistribution of income and wealth.

LOS 14.b: Describe roles and objectives of fiscal policy as well as arguments as to whether the size of a national debt relative to GDP matters.

Objectives of fiscal policy may include the following:

- Influencing the level of economic activity and aggregate demand
- Redistributing wealth and income among segments of the population
- Allocating resources among economic agents and sectors in the economy

In general, decreased taxes and increased government spending both *increase* a budget deficit, overall demand, economic growth, and employment; increased taxes and decreased government spending *decrease* these. Budget deficits are increased in

response to recessions, and budget deficits are decreased to slow growth when inflation is too high.

Keynesian economists believe that fiscal policy, through its effect on aggregate demand, can have a strong effect on economic growth when the economy is operating at less than full employment. Monetarists believe that the effect of fiscal stimulus is only temporary, and that monetary policy should be used to increase or decrease inflationary pressures over time. Monetarists do not believe that monetary policy should be used in an attempt to influence aggregate demand to counter cyclical movements in the economy.

Discretionary fiscal policy refers to the spending and taxing decisions of a national government that are intended to stabilize the economy. In contrast, **automatic stabilizers** are built-in fiscal devices triggered by the state of the economy. For example, during a recession, tax receipts will fall, and government expenditures on unemployment insurance payments will increase. Both of these tend to increase budget deficits and are expansionary. Similarly, during boom times, higher tax revenues coupled with lower outflows for social programs tend to decrease budget deficits and are contractionary.

When a government runs fiscal deficits, it incurs debt that needs to be repaid as well as ongoing interest expense. Total deficits, annual deficits, and interest expense can all be evaluated relative to annual GDP. When these ratios increase beyond certain levels, it may be a cause for concern, and the solvency of the country may be questioned.

A country's **debt ratio** is the ratio of aggregate debt to GDP. Because taxes are linked to GDP, when an economy grows in real terms, so will tax revenues. If the real interest rate on the government's debt is higher than the real growth rate of the economy, then the debt ratio will increase over time (keeping tax rates constant). Similarly, if the real interest rate on the government's debt is lower than real growth in GDP, the debt ratio will decrease (i.e., improve) over time.

Arguments *for* being concerned with the size of a fiscal deficit are as follows:

- Higher deficits lead to higher future taxes. Higher future taxes will lead to disincentives to work and entrepreneurship. This leads to lower long-term economic growth.
- If markets lose confidence in the government, investors may not be willing to refinance the debt. This can lead to the government defaulting (if debt is in a foreign currency) or having to simply print money (if the debt is in local currency). Printing money would ultimately lead to higher inflation.
- Increased government borrowing will tend to increase interest rates, and firms may reduce their borrowing and investment spending as a result, thus decreasing the impact on aggregate demand of deficit spending. This is referred to as the **crowding-out effect** because government borrowing is taking the place of private-sector borrowing.

Arguments *against* being concerned with the size of a fiscal deficit are as follows:

- If the debt is primarily held by domestic citizens, the scale of the problem is overstated.
- If the debt is used to finance productive capital investment, future economic gains will be sufficient to repay the debt.
- Fiscal deficits may prompt needed tax reform.
- Deficits would not matter if **Ricardian equivalence** holds, which means private-sector savings in anticipation of future tax liabilities just offset the government deficit.
- If the economy is operating at less than full capacity, deficits do not divert capital away from productive uses and can aid in increasing GDP and employment.



MODULE QUIZ 14.1

1. Both monetary and fiscal policy are used to:
   A. balance the budget.
   B. achieve economic targets.
   C. redistribute income and wealth.
2. Roles and objectives of fiscal policy *most likely* include:
   A. controlling the money supply to limit inflation.
   B. adjusting tax rates to influence aggregate demand.
   C. using government spending to control interest rates.

MODULE 14.2: FISCAL POLICY TOOLS AND IMPLEMENTATION



Video covering this content is available online.

LOS 14.c: Describe tools of fiscal policy, including their advantages and disadvantages.

Fiscal policy tools include spending tools and revenue tools.

Spending Tools

Transfer payments, also known as entitlement programs, redistribute wealth, taxing some and making payments to others. Examples include government-run retirement income plans (such as Social Security in the United States) and unemployment insurance benefits. Transfer payments are not included in GDP computations.

Current spending refers to government purchases of goods and services on an ongoing and routine basis.

Capital spending refers to government spending on infrastructure, such as roads, schools, bridges, and hospitals. Capital spending is expected to boost future productivity of the economy.

Spending tools are commonly justified as means of pursuing the following goals:

- Provide services such as national defense that benefit all the residents in a country.
- Invest in infrastructure to enhance economic growth.

- Support the country's growth and unemployment targets by directly affecting aggregate demand.
- Provide a minimum standard of living.
- Subsidize investment in research and development for certain high-risk ventures consistent with future economic growth or other goals (e.g., green technology).

Revenue Tools

Direct taxes are levied on income or wealth. These include income taxes, taxes on income for national insurance, wealth taxes, estate taxes, corporate taxes, capital gains taxes, and Social Security taxes. Some progressive taxes (such as income taxes and wealth taxes) collect revenue for wealth and income redistributing.

Indirect taxes are levied on goods and services. These include sales taxes, value-added taxes (VATs), and excise taxes. Indirect taxes can be used to reduce consumption of some goods and services (e.g., alcohol, tobacco, gambling).

Desirable attributes of tax policy are as follows:

- *Simplicity* to use and enforce
- *Efficiency*, defined here as minimizing interference with market forces and not acting as a deterrent to working
- *Fairness* is quite subjective, but two of the commonly held beliefs are **horizontal equality** (people in similar situations should pay similar taxes) and **vertical equality** (richer people should pay more in taxes)
- *Sufficiency*, in that taxes should generate enough revenues to meet the spending needs of the government

Advantages of fiscal policy tools include:

- Social policies (e.g., discouraging tobacco use) can be implemented quickly via indirect taxes.
- Quick implementation of indirect taxes also means that government revenues can be increased without significant additional costs.

Disadvantages of fiscal policy tools include:

- Direct taxes and transfer payments take time to implement, delaying the impact of fiscal policy.
- Capital spending also takes a long time to implement; the economy may have recovered by the time its impact is felt.

Announcing a change in fiscal policy may have significant effects on expectations. For example, an announcement of a future increase in taxes may immediately reduce current consumption, rapidly producing the desired goal of reducing aggregate demand.

Not all fiscal policy tools affect economic activity equally. Spending tools are most effective in increasing aggregate demand. Tax reductions are somewhat less effective, as people may not spend the entire amount of the tax savings. Tax reductions for those with low incomes will be more effective in increasing aggregate demand, as those with

lower incomes tend to spend a larger proportion of income on consumption; that is, they save a smaller proportion of income and have a higher **marginal propensity to consume (MPC)**.

Fiscal Multiplier

Changes in government spending have magnified effects on aggregate demand because those whose incomes increase from increased government spending will, in turn, increase their spending, which increases the incomes and spending of others. The magnitude of the *multiplier effect* depends on the tax rate and on the marginal propensity to consume.

To understand the calculation of the multiplier effect, consider an increase in government spending of $100 when the MPC is 80% and the tax rate is 25%. The increase in spending increases incomes by $100, but $25 (100 × 0.25) of that will be paid in taxes. **Disposable income** is equal to income after taxes, so disposable income increases by $100 × (1 − 0.25) = $75. With an MPC of 80%, additional spending by those who receive the original $100 increase is $75 × 0.8 = $60.

This additional spending will increase others' incomes by $60 and disposable incomes by $60 × 0.75 = $45, from which they will spend $45 × 0.8 = $36.

Because each iteration of this process reduces the amount of additional spending, the effect reaches a limit. The **fiscal multiplier** determines the potential increase in aggregate demand resulting from an increase in government spending:

$$\text{fiscal multiplier} = \frac{1}{1 - \text{MPC}(1 - t)}$$

Here, with a tax rate of 25% and an MPC of 80%, the fiscal multiplier is 1 / [1 − 0.8(1 − 0.25)] = 2.5, and the increase of $100 in government spending has the potential to increase aggregate demand by $250.

The fiscal multiplier is inversely related to the tax rate (higher tax rate decreases the multiplier) and is directly related to the marginal propensity to consume (higher MPC increases the multiplier).

Balanced Budget Multiplier

To balance the budget, the government could increase taxes by $100 to just offset a $100 increase in spending. Changes in taxes also have a magnified effect on aggregate demand. An increase in taxes will decrease disposable income and consumption expenditures, thereby decreasing aggregate demand. The initial decrease in spending from a tax increase of $100 is 100 × MPC = 100 × 0.8 = $80; beyond that, the multiplier effect is the same as we described for a direct increase in government spending, and the overall decrease in aggregate demand for a $100 tax increase is 100(MPC) × fiscal multiplier—or, for our example, 100(0.8)(2.5) = $200.

Combining the total increase in aggregate demand from a $100 increase in government spending with the total decrease in aggregate demand from a $100 tax increase shows

that the net effect on aggregate demand of both is an increase of $250 – $200 = $50, so we can say that the balanced budget multiplier is positive.

If, instead of a $100 increase in taxes, we increased taxes by 100 / MPC = 100 / 0.8 = $125 and increased government spending by $100, the net effect on aggregate demand would be zero.

Ricardian Equivalence

Increases in the current deficit mean greater taxes in the future. To maintain their preferred pattern of consumption over time, taxpayers may increase current savings (reduce current consumption) to offset the expected cost of higher future taxes. If taxpayers reduce current consumption and increase current saving by just enough to repay the principal and interest on the debt the government issued to fund the increased deficit, there is no effect on aggregate demand. This is known as Ricardian equivalence, after economist David Ricardo. If taxpayers underestimate their future liability for servicing and repaying the debt so that aggregate demand is increased by equal spending and tax increases, Ricardian equivalence does not hold. Whether it does is an open question.

LOS 14.d: Explain the implementation of fiscal policy and difficulties of implementation as well as whether a fiscal policy is expansionary or contractionary.

Fiscal policy is implemented through changes in taxes and spending. This is called **discretionary fiscal policy** (as opposed to automatic stabilizers, discussed previously). Discretionary fiscal policy would be designed to be expansionary when the economy is operating below full employment. Fiscal policy aims to stabilize aggregate demand. During recessions, actions can be taken to increase government spending or decrease taxes. Either change tends to strengthen the economy by increasing aggregate demand, putting more money in the hands of corporations and consumers to invest and spend. During inflationary economic booms, actions can be taken to decrease government spending or increase taxes. Either change tends to slow the economy by decreasing aggregate demand, taking money out of the hands of corporations and consumers and causing both investment and consumption spending to decrease.

Discretionary fiscal policy is not an exact science. First, economic forecasts might be wrong, leading to incorrect policy decisions. Second, complications arise in practice that delay both the implementation of discretionary fiscal policy and the impact of policy changes on the economy. The lag between recessionary or inflationary conditions in the economy and the impact on the economy of fiscal policy changes can be divided into three types:

1. **Recognition lag.** Discretionary fiscal policy decisions are made by a political process. The state of the economy is complex, and it may take policymakers time to recognize the nature and extent of the economic problems.
2. **Action lag.** Governments take time to discuss, vote on, and enact fiscal policy changes.

3. **Impact lag.** Time elapses between the enactment of fiscal policy changes and when the impact of the changes on the economy actually takes place. It takes time for corporations and individuals to act on the fiscal policy changes, and fiscal multiplier effects occur only over time as well.

These lags can actually make fiscal policy counterproductive. For example, if the economy is in a recession phase, fiscal stimulus may be deemed appropriate. However, by the time fiscal stimulus is implemented and has its full impact, the economy may already be on a path to a recovery driven by the private sector.

Additional macroeconomic issues may hinder the usefulness of fiscal policy:

- *Misreading economic statistics*. The full employment level for an economy is not precisely measurable. If the government relies on expansionary fiscal policy mistakenly at a time when the economy is already at full capacity, it will simply drive inflation higher.
- *Crowding-out effect*. Expansionary fiscal policy may crowd out private investment. Greater government borrowing tends to increase interest rates, which decreases private investments (some investments will be uneconomic with higher borrowing costs). This crowding-out effect can reduce the impact of expansionary fiscal policy on aggregate demand. Opinions vary on the magnitude of this effect.
- *Supply shortages*. If economic activity is slow due to resource constraints (low availability of labor or other resources) and not due to low demand, expansionary fiscal policy will fail to achieve its objective and will probably lead to higher inflation.
- *Limits to deficits*. There is a limit to expansionary fiscal policy. If the markets perceive that the deficit is already too high as a proportion of GDP, funding the deficit will be problematic. This could lead to higher interest rates and actually make the situation worse.
- *Multiple targets*. If the economy has high unemployment coupled with high inflation, fiscal policy cannot address both problems simultaneously.

To determine if fiscal policy is expansionary or contractionary, economists often focus on *changes* in the budget surplus or deficit. An increase (decrease) in surplus is indicative of a contractionary (expansionary) fiscal policy. Similarly, an increase (decrease) in deficit is indicative of an expansionary (contractionary) fiscal policy.



PROFESSOR'S NOTE

For the exam, an increase (decrease) in a revenue item (e.g., sales tax) should be considered contractionary (expansionary), and an increase (decrease) in a spending item (e.g., construction of highways) should be considered expansionary (contractionary).

A government's intended fiscal policy is not necessarily obvious from just examining the current deficit. Consider an economy that is in recession so that transfer payments are increased and tax revenue is decreased, leading to a deficit. This does not necessarily indicate that fiscal policy is expansionary as, at least to some extent, the deficit is a natural outcome of the recession without any explicit action of the

government. Economists often use a measure called the **structural budget deficit** (or **cyclically adjusted budget deficit**) to gauge fiscal policy. This is the deficit that would occur based on current policies if the economy were at full employment.



MODULE QUIZ 14.2

1. A government enacts a program to subsidize farmers with an expansive spending program of $10 billion. At the same time, the government enacts a $10 billion tax increase over the same period. Which of the following statements *best* describes the impact on aggregate demand?
   A. Lower growth, because the tax increase will have a greater effect.
   B. No effect, because the tax and spending effects just offset each other.
   C. Higher growth, because the spending increase will have a greater effect.
2. A government reduces spending by $50 million. The tax rate is 30%, and consumers exhibit a marginal propensity to consume of 80%. The change in aggregate demand caused by the change in government spending is *closest* to:
   A. −$66 million.
   B. −$114 million.
   C. −$250 million.
3. The size of a national debt is *most likely* to be a concern for policymakers if:
   A. Ricardian equivalence holds.
   B. a crowding-out effect occurs.
   C. debt is used to finance capital growth.
4. A government is concerned about the timing of fiscal policy changes and is considering requiring the compilation and reporting of economic statistics weekly, rather than quarterly. The new reporting frequency is intended to decrease the:
   A. action lag.
   B. impact lag.
   C. recognition lag.
5. Fiscal policy is *most likely* to be expansionary if tax rates:
   A. and government spending both decrease.
   B. decrease and government spending increases.
   C. increase and government spending decreases.

KEY CONCEPTS

LOS 14.a

Fiscal policy is a government's use of taxation and spending to influence the economy. Monetary policy deals with determining the quantity of money supplied by the central bank. Both policies aim to achieve economic growth with price level stability, although governments use fiscal policy for social and political reasons as well.

LOS 14.b

Objectives of fiscal policy can include (1) influencing the level of economic activity, (2) redistributing wealth or income, and (3) allocating resources among industries.

There are arguments for being concerned with the size of a fiscal deficit:

- Higher future taxes lead to disincentives to work, negatively affecting long-term economic growth.

- Fiscal deficits may not be financed by the market when debt levels are high.
- There is a crowding-out effect as government borrowing increases interest rates and decreases the private-sector investment.

There are arguments against being concerned with the size of a fiscal deficit:

- Debt may be financed by domestic citizens.
- Deficits for capital spending can boost the productive capacity of the economy.
- Fiscal deficits may prompt needed tax reform.
- Ricardian equivalence may prevail: private savings rise in anticipation of the need to repay principal on government debt.
- When the economy is operating below full employment, deficits do not crowd out private investment.

LOS 14.c

Fiscal policy tools include spending tools and revenue tools. Spending tools include transfer payments, current spending (goods and services used by government), and capital spending (investment projects funded by government). Revenue tools include direct and indirect taxation.

An advantage of fiscal policy is that indirect taxes can be used to quickly implement social policies and can also be used to quickly raise revenues at a low cost.

Disadvantages of fiscal policy include time lags for implementing changes in direct taxes and time lags for capital spending changes to have an impact.

LOS 14.d

Fiscal policy is implemented by government changes in taxing and spending policies. Delays in realizing the effects of fiscal policy changes limit their usefulness. Delays can be caused by the following:

- *Recognition lag*. Policymakers may not immediately recognize when fiscal policy changes are needed.
- *Action lag*. Governments take time to enact needed fiscal policy changes.
- *Impact lag*. Fiscal policy changes take time to affect economic activity.

In general, fiscal policies that increase the government deficit are often considered expansionary, and policies that decrease the deficit are considered contractionary. However, because a deficit will naturally increase during a recession and decrease during an expansion, the impact of a current deficit is often judged relative to a structural (cyclically adjusted) deficit amount.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 14.1

1. **B** Both monetary and fiscal policies primarily strive to achieve economic targets such as inflation and GDP growth. Balancing the budget is not a goal for monetary policy and is a potential outcome of fiscal policy. Fiscal policy (but not monetary policy) may secondarily be used as a tool to redistribute income and wealth. (LOS 14.a)

2. **B** Influencing the level of aggregate demand through taxation and government spending is an objective of fiscal policy. Controlling inflation and interest rates are typical objectives of monetary policy. (LOS 14.b)

Module Quiz 14.2

1. **C** The amount of the spending program exactly offsets the amount of the tax increase, leaving the budget unaffected. The multiplier for government spending is greater than the multiplier for a tax increase. Therefore, the balanced budget multiplier is positive. All of the government spending enters the economy as increased expenditure, whereas spending is reduced by only a portion of the tax increase. (LOS 14.c)

2. **B** fiscal multiplier = 1 / [1 − MPC (1 − T)] = 1 / [1 − 0.80(1 − 0.3)] = 2.27

   change in government spending = −$50 million

   change in aggregate demand = −(50 × 2.27) = −$113.64 million

   (LOS 14.c)

3. **B** Crowding out refers to the possibility that government borrowing causes interest rates to increase and private investment to decrease. If government debt is financing the growth of productive capital, this should increase future economic growth and tax receipts to repay the debt. Ricardian equivalence is the theory that if government debt increases, private citizens will increase savings in anticipation of higher future taxes, and it is an argument against being concerned about the size of government debt and budget deficits. (LOS 14.d)

4. **C** More frequent and current economic data would make it easier for authorities to monitor the economy and to recognize problems. The reduction in the time between economic reports should reduce the recognition lag. (LOS 14.d)

5. **B** Increases in government spending and decreases in taxes are expansionary fiscal policy. Decreases in spending and increases in taxes are contractionary fiscal policy. (LOS 14.d)

READING 15

MONETARY POLICY

MODULE 15.1: CENTRAL BANK OBJECTIVES AND TOOLS



Video covering this content is available online.

LOS 15.a: Describe the roles and objectives of central banks.

There are several key **roles of central banks**:

1. *Sole supplier of currency*. Central banks have the sole authority to supply money. Traditionally, such money was backed by gold; the central bank stood ready to convert the money into a prespecified quantity of gold. Later on, the gold backing was removed, and money supplied by the central bank was deemed **legal tender** by law. Money not backed by any tangible value is termed **fiat money**. As long as fiat money holds its value over time and is acceptable for transactions, it can continue to serve as a medium of exchange.
2. *Banker to the government and other banks*. Central banks provide banking services to the government and other banks in the economy.
3. *Regulator and supervisor of payments system*. In many countries, central banks may regulate the banking system by imposing standards of risk-taking allowed and reserve requirements of banks under its jurisdiction. Central banks also oversee the payments system to ensure smooth clearing operations domestically and in conjunction with other central banks for international transactions.
4. *Lender of last resort*. Central banks' ability to print money allows them to supply money to banks that are experiencing shortages. This government backing tends to prevent runs on banks (i.e., large-scale withdrawals) by assuring depositors that their funds are secure.
5. *Holder of gold and foreign exchange reserves*. Central banks are often the repositories of the nation's gold and reserves of foreign currencies.
6. *Conductor of monetary policy*. Central banks control or influence the quantity of money supplied in an economy and growth of money supply over time.

The primary objective of a central bank is to *control inflation* so as to promote price stability. High inflation is not conducive to a stable economic environment. High inflation leads to **menu costs** (i.e., cost to businesses of constantly having to change their prices) and **shoe leather costs** (i.e., costs to individuals of making frequent trips

to the bank so as to minimize their holdings of cash that are depreciating in value due to inflation).

In addition to price stability, some central banks have other stated goals:

- Stability in exchange rates with foreign currencies
- Full employment
- Sustainable positive economic growth
- Moderate long-term interest rates

The target inflation rate in most developed countries is a range around 2% to 3%. A target of zero inflation is not used because that increases the risk of deflation, which can be disruptive for an economy.

While most developed countries have an explicit target inflation rate, the U.S. Fed and the Bank of Japan do not. In the United States, this is because the Fed has the additional goals of maximum employment and moderate long-term interest rates. In Japan, it is because deflation, rather than inflation, has been a persistent problem in recent years.

Some developed countries, and several developing countries, choose a target level for the exchange rate of their currency with that of another country, primarily the U.S. dollar. This is referred to as **pegging** their exchange rate with the dollar. If their currency appreciates (i.e., becomes relatively more valuable), they can sell their domestic currency reserves for dollars to reduce the exchange rate. While such actions may be effective in the short run, for stability of the exchange rate over time, the monetary authorities in the pegging country must manage interest rates and economic activity to achieve their goal. This can lead to increased volatility of their money supply and interest rates. The pegging country essentially commits to a policy intended to make its inflation rate equal to the inflation rate of the country to which it pegs its currency.

LOS 15.b: Describe tools used to implement monetary policy tools and the monetary transmission mechanism, and explain the relationships between monetary policy and economic growth, inflation, interest, and exchange rates.

Monetary policy is implemented using the **monetary policy tools** of the central bank. The three main policy tools of central banks are as follows:

1. *Policy rate*. In the United States, banks can borrow funds from the Fed if they have temporary shortfalls in reserves. The rate at which banks can borrow reserves from the Fed is called the *discount rate*. For the European Central Bank (ECB), it is called the *refinancing rate*.

   One way to lend money to banks is through a *repurchase agreement*. The central bank purchases securities from banks that, in turn, agree to repurchase the securities at a higher price in the future. The percentage difference between the purchase price and the repurchase price is effectively the rate at which the central bank is lending to member banks. The Bank of England uses this method, and its policy rate is called the *two-week repo (repurchase) rate*. A lower rate reduces banks' cost of funds,

encourages lending, and tends to decrease interest rates overall. A higher policy rate has the opposite effect, decreasing lending and increasing interest rates.

In the United States, the *federal funds rate* is the rate that banks charge each other on overnight loans of reserves. The Fed sets a target for this market-determined rate and uses open market operations to influence it toward the target rate.

2. *Reserve requirements*. By increasing the reserve requirement (the percentage of deposits that banks are required to retain as reserves), the central bank effectively decreases the funds that are available for lending and thereby decreases the money supply, which will tend to increase interest rates. A decrease in the reserve requirement will increase the funds available for lending and the money supply, which will tend to decrease interest rates. This tool only works well to increase the money supply if banks are willing to lend and customers are willing to borrow.
3. *Open market operations*. Buying and selling of securities by the central bank is referred to as open market operations. When the central bank buys securities, cash replaces securities in investor accounts, banks have excess reserves, more funds are available for lending, the money supply increases, and interest rates decrease. Sales of securities by the central bank have the opposite effect, reducing cash in investor accounts, excess reserves, funds available for lending, and the money supply, which will tend to cause interest rates to increase. In the United States, open market operations are the Fed's most commonly used tool and are important in achieving the federal funds target rate.

Monetary Transmission Mechanism

The **monetary transmission mechanism** refers to the ways in which a change in monetary policy, specifically the central bank's policy rate, affects the price level and inflation. A change in the policy rates that the monetary authorities control directly is transmitted to prices through four channels: other short-term rates, asset values, currency exchange rates, and expectations.

We can examine the transmission mechanism in more detail by considering the effects of a change to a contractionary monetary policy implemented through an increase in the policy rate:

- Banks' *short-term lending rates will increase* in line with the increase in the policy rate. The higher rates will decrease aggregate demand as consumers reduce credit purchases and businesses cut back on investment in new projects.
- Bond prices, equity prices, and *asset prices in general will decrease* as the discount rates applied to future expected cash flows are increased. This may have a wealth effect because a decrease in the value of households' assets may increase the savings rate and decrease consumption.
- Both consumers and businesses may decrease their expenditures because their *expectations for future economic growth decrease*.
- The increase in interest rates may attract foreign investment in debt securities, leading to an *appreciation of the domestic currency relative to foreign currencies*. An

appreciation of the domestic currency increases the foreign currency prices of exports and can reduce demand for the country's export goods.

Taken together, these effects act to decrease aggregate demand and put downward pressure on the price level. A decrease in the policy rate would affect the price level through the same channels, but in the opposite direction.

Monetary Policy Relation With Economic Growth, Inflation, Interest, and Exchange Rates

If money neutrality holds, changes in monetary policy and the policy rate will have no effect on real output. In the short run, however, changes in monetary policy can affect real economic growth as well as interest rates, inflation, and foreign exchange rates. The effects of a change to a more expansionary monetary policy may include any or all of the following:

- The central bank buys securities, which increases bank reserves.
- The interbank lending rate decreases as banks are more willing to lend each other reserves.
- Other short-term rates decrease as the increase in the supply of loanable funds decreases the equilibrium rate for loans.
- Longer-term interest rates also decrease.
- The decrease in real interest rates causes the currency to depreciate in the foreign exchange market.
- The decrease in long-term interest rates increases business investments in plant and equipment.
- Lower interest rates cause consumers to increase their purchases of houses, autos, and durable goods.
- Depreciation of the currency increases foreign demand for domestic goods.
- The increases in consumption, investment, and net exports all increase aggregate demand.
- The increase in aggregate demand increases inflation, employment, and real GDP.



MODULE QUIZ 15.1

1. A central bank's policy goals *least likely* include:
   A. price stability.
   B. minimizing long-term interest rates.
   C. maximizing the sustainable growth rate of the economy.
2. A country that targets a stable exchange rate with another country's currency *least likely*:
   A. accepts the inflation rate of the other country.
   B. will sell its currency if its foreign exchange value rises.
   C. must also match the money supply growth rate of the other country.
3. If a country's inflation rate is below the central bank's target rate, the central bank is *most likely* to:
   A. sell government securities.
   B. increase the reserve requirement.

C. decrease the overnight lending rate.

4. A central bank conducts monetary policy primarily by altering the:
   A. policy rate.
   B. inflation rate.
   C. long-term interest rate.
5. Purchases of securities in the open market by the monetary authorities are *least likely* to increase:
   A. excess reserves.
   B. cash in investor accounts.
   C. the interbank lending rate.
6. An increase in the policy rate will *most likely* lead to an increase in:
   A. business investments in fixed assets.
   B. consumer spending on durable goods.
   C. the foreign exchange value of the domestic currency.

MODULE 15.2: MONETARY POLICY EFFECTS AND LIMITATIONS



Video covering this content is available online.

LOS 15.c: Describe qualities of effective central banks; contrast their use of inflation, interest rate, and exchange rate targeting in expansionary or contractionary monetary policy; and describe the limitations of monetary policy.

For a central bank to succeed in its inflation targeting policies, it should have three essential qualities:

1. *Independence*. For a central bank to be effective in achieving its goals, it should be free from political interference. Reducing the money supply to reduce inflation can also be expected to decrease economic growth and employment. The political party in power has an incentive to boost economic activity and reduce unemployment before elections. For this reason, politicians may interfere with the central bank's activities, compromising its ability to manage inflation. Independence should be thought of in relative terms (degrees of independence) rather than absolute terms. Even in the case of relatively independent central banks, the heads of the banks may be appointed by politicians.

   Independence can be evaluated based on both **operational independence** and **target independence**. Operational independence means that the central bank is allowed to independently determine the policy rate. Target independence means the central bank also defines how inflation is computed, sets the target inflation level, and determines the horizon over which the target is to be achieved. The ECB has both target and operational independence, while most other central banks have only operational independence.
2. *Credibility*. To be effective, central banks should follow through on their stated intentions. If a government with large debts, instead of a central bank, sets an inflation target, the target would not be credible because the government has an incentive to allow inflation to exceed the target level. On the other hand, a credible central bank's targets can become self-fulfilling prophecies. If the market believes

that a central bank is serious about achieving a target inflation rate of 3%, wages and other nominal contracts will be based on 3% inflation, and actual inflation will then be close to that level.

3. *Transparency*. Transparency on the part of central banks aids their credibility. Transparency means that central banks periodically disclose the state of the economic environment by issuing **inflation reports**. Transparent central banks periodically report their views on the economic indicators and other factors they consider in their interest rate setting policy. When a central bank makes clear the economic indicators it uses in establishing monetary policy and how they will be used, it not only gains credibility, but also makes policy changes easier to anticipate and implement.

Inflation, Interest Rate, and Exchange Rate Targeting by Central Banks

Central banks have used various economic variables and indicators over the years to make monetary policy decisions. In the past, some have used **interest rate targeting**, increasing the money supply when specific interest rates rose above the target band and decreasing the money supply (or the rate of money supply growth) when interest rates fell below the target band. Currently, **inflation targeting** is the most widely used tool for making monetary policy decisions, and it is the method required by law in some countries. Central banks that currently use inflation targeting include the United Kingdom, Brazil, Canada, Australia, Mexico, and the ECB.

The **neutral interest rate** of an economy is the growth rate of the money supply that neither increases nor decreases the economic growth rate:

$$\text{neutral interest rate} = \text{real trend rate of economic growth} + \text{inflation target}$$

When the policy rate is above the neutral rate, monetary policy is said to be contractionary; when the policy rate is below the neutral rate, monetary policy is said to be expansionary.

The most common inflation rate target is 2%, with a permitted deviation of ±1% so the target band is 1% to 3%. The reason the inflation target is not 0% is that variations around that rate would allow for negative inflation (i.e., deflation), which is considered disruptive to the smooth functioning of an economy. Central banks are not necessarily targeting current inflation, which is the result of prior policy and events, but inflation in the range of two years in the future.

Some countries, especially developing countries, use **exchange rate targeting**. That is, they target a foreign exchange rate between their currency and another (often the U.S. dollar), rather than targeting inflation. As an example, consider a country that has targeted an exchange rate for its currency versus the U.S. dollar. If the foreign exchange value of the domestic currency falls relative to the U.S. dollar, the monetary authority must use foreign reserves to purchase their domestic currency (which will reduce money supply growth and increase interest rates) to reach the target exchange rate. Conversely, an increase in the foreign exchange value of the domestic currency above the target rate will require sale of the domestic currency in currency markets to reduce

its value (increasing the domestic money supply and decreasing interest rates) to move toward the target exchange rate. One result of exchange rate targeting may be greater volatility of the money supply because domestic monetary policy must adapt to the necessity of maintaining a stable foreign exchange rate.

Over the short term, the targeting country can purchase or sell its currency in the foreign exchange markets to influence the exchange rate. There are limits, however, on how much influence currency purchases or sales can have on exchange rates over time. For example, a country may run out of foreign reserves with which to purchase its currency when the exchange value of its currency is still below the target exchange rate.

The net effect of exchange rate targeting is that the targeting country will have the same inflation rate as the country with the targeted currency, and the targeting country will need to follow monetary policy and accept interest rates that are consistent with this goal regardless of domestic economic circumstances.

Limitations of Monetary Policy

This transmission mechanism for monetary policy previously described does not always produce the intended results. In particular, long-term rates may not rise and fall with short-term rates because of the effect of monetary policy changes on expected inflation.

If individuals and businesses believe that a decrease in the money supply intended to reduce inflation will be successful, they will expect lower future inflation rates. Because long-term bond yields include a premium for expected inflation, long-term rates could fall (tending to increase economic growth), even while the central bank has increased short-term rates to slow economic activity. Conversely, increasing the money supply to stimulate economic activity could lead to an increase in expected inflation rates and long-term bond yields, even as short-term rates fall.

From a different perspective, monetary tightening may be viewed as too extreme—increasing the probability of a recession, making long-term bonds more attractive, and reducing long-term interest rates. If money supply growth is seen as inflationary, higher expected future asset prices will make long-term bonds relatively less attractive and will increase long-term interest rates. Bond market participants that act in this way have been called **bond market vigilantes**. When the central bank's policy is credible and investors believe that the inflation target rate will be maintained over time, this effect on long-term rates will be small.

Another situation in which the transmission mechanism may not perform as expected is if demand for money becomes very elastic and individuals willingly hold more money, even without a decrease in short-term rates. Such a situation is called a **liquidity trap**. Increasing the growth of the money supply will not decrease short-term rates under these conditions because individuals hold the money in cash balances instead of investing in interest-bearing securities. If an economy is experiencing deflation even though monetary policy has been expansionary, liquidity trap conditions may be present.

Compared to inflation, deflation is more difficult for central banks to reverse. In a deflationary environment, monetary policy needs to be expansionary. However, the central bank is limited to reducing the nominal policy rate to zero. Once it reaches zero, the central bank has a limited ability to further stimulate the economy.

Another reason standard tools for increasing the money supply might not increase economic activity is that even with increasing excess reserves, banks may not be willing to lend. When what has become known as the *credit bubble* collapsed in 2008, banks around the world lost equity capital and desired to rebuild it. For this reason, they decreased their lending even as money supplies were increased, and short-term rates fell. With short-term rates near zero, economic growth still poor, and a threat of deflation, central banks began a policy termed **quantitative easing**.

In the United Kingdom, quantitative easing entailed large purchases of British government bonds in the maturity range of three to five years. The intent was to reduce interest rates to encourage borrowing and to generate excess reserves in the banking system to encourage lending. Uncertainty about the economy's future caused banks to behave quite conservatively and willingly hold more excess reserves, rather than make loans.

In the United States, billions of dollars were made available for the Fed to buy assets other than short-term Treasury securities. Large amounts of mortgage securities were purchased from banks to encourage bank lending and to reduce mortgage rates in an attempt to revive the housing market, which had collapsed. When this program did not have the desired effect, a second round of quantitative easing (QE2) was initiated. The Fed purchased long-term Treasury bonds in large quantities (hundreds of billions of dollars) with the goal of bringing down longer-term interest rates and generating excess reserves to increase lending and economic growth. The Fed has also purchased securities with credit risk as part of its quantitative easing, improving banks' balance sheets but perhaps shifting risk from the private sector to the public sector.

Monetary Policy in Developing Economies

Developing countries face problems in successfully implementing monetary policy. Without a liquid market in their government debt, interest rate information may be distorted and open market operations difficult to implement. In a very rapidly developing economy, it may be quite difficult to determine the neutral rate of interest for policy purposes. Rapid financial innovation may change the demand to hold monetary aggregates. Central banks may lack credibility because of past failure(s) to maintain inflation rates in a target band and might not be given independence by the political authority.

LOS 15.d: Explain the interaction of monetary and fiscal policy.

Monetary policy and fiscal policy may each be either expansionary or contractionary, so there are four possible scenarios:

1. **Expansionary fiscal and monetary policy.** In this case, the impact will be highly expansionary, taken together. Interest rates will usually be lower (due to monetary

policy), and the private and public sectors will both expand.

2. **Contractionary fiscal and monetary policy.** In this case, aggregate demand and GDP would be lower, and interest rates would be higher due to tight monetary policy. Both the private and public sectors would contract.
3. **Expansionary fiscal policy and contractionary monetary policy.** In this case, aggregate demand will likely be higher (due to fiscal policy), while interest rates will be higher (due to increased government borrowing and tight monetary policy). Government spending as a proportion of GDP will increase.
4. **Contractionary fiscal policy and expansionary monetary policy.** In this case, interest rates will fall from decreased government borrowing and from the expansion of the money supply, increasing both private consumption and output. Government spending as a proportion of GDP will decrease due to contractionary fiscal policy. The private sector would grow as a result of lower interest rates.

Not surprisingly, the fiscal multipliers for different types of fiscal stimulus differ, and the effects of expansionary fiscal policy are greater when it is combined with expansionary monetary policy. The fiscal multiplier for direct government spending increases has been much higher than the fiscal multiplier for increases in transfers to individuals or tax reductions for workers. Transfer payments to the poor have the greatest relative impact, followed by tax cuts for workers, and broader-based transfers to individuals (not targeted). For all types of fiscal stimulus, the impact is greater when the fiscal actions are combined with expansionary monetary policy. This may reflect the impact of greater inflation, falling real interest rates, and the resulting increase in business investments.



MODULE QUIZ 15.2

1. Qualities of effective central banks include:
   A. credibility and verifiability.
   B. comparability and relevance.
   C. independence and transparency.

2. Monetary policy is likely to be *least* responsive to domestic economic conditions if policymakers employ:
   A. inflation targeting.
   B. interest rate targeting.
   C. exchange rate targeting.

3. Monetary policy is *most likely* to fail to achieve its objectives when the economy is:
   A. growing rapidly.
   B. experiencing deflation.
   C. experiencing disinflation.

KEY CONCEPTS

LOS 15.a

Central bank roles include supplying currency, acting as a banker to the government and to other banks, regulating and supervising the payments system, acting as a lender

of last resort, holding the nation's gold and foreign currency reserves, and conducting monetary policy.

Central banks have the objective of controlling inflation, and some have additional goals of maintaining currency stability, full employment, positive sustainable economic growth, or moderate interest rates.

LOS 15.b

Policy tools available to central banks include the policy rate, reserve requirements, and open market operations. The policy rate is called the discount rate in the United States, the refinancing rate by the ECB, and the two-week repo rate in the United Kingdom.

Decreasing the policy rate, decreasing reserve requirements, and making open market purchases of securities are all expansionary. Increasing the policy rate, increasing reserve requirements, and making open market sales of securities are all contractionary.

The transmission mechanism for changes in the central bank's policy rate through to prices and inflation includes one or more of the following:

- Short-term bank lending rates
- Asset prices
- Expectations for economic activity and future policy rate changes
- Exchange rates with foreign currencies

A contractionary monetary policy (increase in policy rate) will tend to decrease economic growth, increase market interest rates, decrease inflation, and lead to appreciation of the domestic currency in foreign exchange markets. An expansionary monetary policy (decrease in policy rate) will have opposite effects, tending to increase economic growth, decrease market interest rates, increase inflation, and reduce the value of the currency in foreign exchange markets.

LOS 15.c

Effective central banks exhibit independence, credibility, and transparency.

- *Independence*. The central bank is free from political interference.
- *Credibility*. The central bank follows through on its stated policy intentions.
- *Transparency*. The central bank makes it clear what economic indicators it uses and reports on the state of those indicators.

Most central banks set target inflation rates, typically 2%–3%, rather than targeting interest rates, as was once common. When inflation is expected to rise above (fall below) the target band, the money supply is decreased (increased) to reduce (increase) economic activity.

Developing economies sometimes target a stable exchange rate for their currency relative to that of a developed economy, selling their currency when its value rises above the target rate and buying their currency with foreign reserves when the rate falls below the target. The developing country must follow a monetary policy that

supports the target exchange rate and essentially commits to having the same inflation rate as the developed country.

Reasons that monetary policy may not work as intended:

- Monetary policy changes may affect inflation expectations to such an extent that long-term interest rates move opposite to short-term interest rates.
- Individuals may be willing to hold greater cash balances without a change in short-term rates (liquidity trap).
- Banks may be unwilling to lend greater amounts, even when they have increased excess reserves.
- Short-term rates cannot be reduced below zero.
- Developing economies face unique challenges in using monetary policy due to undeveloped financial markets, rapid financial innovation, and lack of credibility of the monetary authority.

LOS 15.d

Interaction of monetary and fiscal policies:

Monetary Policy	Fiscal Policy	Interest Rates	Output	Private Sector Spending	Public Sector Spending
Tight	Tight	Higher	Lower	Lower	Lower
Easy	Easy	Lower	Higher	Higher	Higher
Tight	Easy	Higher	Higher	Lower	Higher
Easy	Tight	Lower	Varies	Higher	Lower

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 15.1

1. **B** Central bank goals often include the following: maximum employment, which is interpreted as the maximum sustainable growth rate of the economy; stable prices; and *moderate* (not minimum) long-term interest rates. (LOS 15.a)

2. **C** The money supply growth rate may need to be adjusted to keep the exchange rate within acceptable bounds, but it is not necessarily the same as that of the other country. The other two statements are true. (LOS 15.a)

3. **C** Decreasing the overnight lending rate would add reserves to the banking system, which would encourage bank lending, expand the money supply, reduce interest rates, and allow GDP growth and the rate of inflation to increase. Selling government securities or increasing the reserve requirement would have the opposite effect, reducing the money supply and decreasing the inflation rate. (LOS 15.b)

4. **A** The primary method by which a central bank conducts monetary policy is through changes in the target short-term rate or policy rate. (LOS 15.b)

5. **C** Open market purchases by monetary authorities decrease the interbank lending rate by increasing excess reserves that banks can lend to one another—and therefore, increasing their willingness to lend. (LOS 15.b)

6. **C** An increase in the policy rate is likely to increase longer-term interest rates, causing decreases in consumption spending on durable goods and business investments in plant and equipment. The increase in rates, however, makes investment in the domestic economy more attractive to foreign investors, increasing demand for the domestic currency and causing the currency to appreciate. (LOS 15.b)

Module Quiz 15.2

1. **C** The three qualities of effective central banks are independence, credibility, and transparency. (LOS 15.c)

2. **C** Exchange rate targeting requires monetary policy to be consistent with the goal of a stable exchange rate with the targeted currency, regardless of domestic economic conditions. (LOS 15.c)

3. **B** Monetary policy has a limited ability to act effectively against deflation because the policy rate cannot be reduced below zero, and demand for money may be highly elastic (liquidity trap). (LOS 15.c)

READING 16

INTRODUCTION TO GEOPOLITICS

MODULE 16.1: GEOPOLITICS



Video covering this content is available online.

LOS 16.a: Describe geopolitics from a cooperation versus competition perspective.

Geopolitics refers to interactions among nations, including the actions of **state actors** (national governments) and **nonstate actors** (corporations, nongovernment organizations, and individuals).

Geopolitics also refers to the study of how geography affects interactions among nations and their citizens. For example, firms located in coastal countries naturally tend to be the dominant participants in international shipping.

One way to examine geopolitics is through analysis of the extent to which individual countries cooperate with one another. Potential areas for cooperation include diplomatic and military matters, and economic and cultural interactions. In terms of economics, areas of cooperation include freedom of movement across borders for goods, services, and capital; agreements to harmonize tariffs; international standardization of rules; and transfers of information and technology.

While a country that engages with other countries on these matters may be considered cooperative and one that does not may be considered noncooperative, the extent of cooperation actually varies along a spectrum. A country might be more cooperative on some issues and less cooperative on others, and its degree of cooperation can change over time or with the outcomes of the country's domestic politics. A country's current decision makers and the length of its political cycle are factors to consider when analyzing geopolitics.

A country will typically cooperate with other countries when doing so advances its national interests. For example, a country may cooperate with its neighbors in a military alliance if doing so will further its interests in protecting its citizens from foreign invaders.

We can analyze a country's national interests as a hierarchy, with its top priorities being those that ensure its survival. A country's **geophysical resource endowment** may influence its priorities. For example, a country that has mineral resources but lacks arable land needs to trade minerals for food; therefore, it has an interest in cooperating with other countries to keep international trade lanes open.

Nonstate actors often have interests in cooperating across borders. Individuals and firms seek to direct their resources to their highest-valued uses, and some of those uses may be in other countries. To facilitate the flow of resources, state and nonstate actors may cooperate on **standardization** of regulations and processes. One key example of standardization among countries is International Financial Reporting Standards for firms presenting their accounting data to the public, which we will examine in the Financial Statement Analysis topic area.

Cultural factors, such as historical emigration patterns or a shared language, can be another influence on a country's level of cooperation. Among these cultural factors are a country's formal and informal **institutions**, such as laws, public and private organizations, or distinct customs and habits. Strong and stable institutions can make cooperation easier for state and nonstate actors. For example, countries that produce and export large amounts of cultural content tend to be those with legal and ethical institutions that protect intellectual property. Cultural exchange is one means through which a country may exercise **soft power**, which is the ability to influence other countries without using or threatening force.

LOS 16.b: Describe geopolitics and its relationship with globalization.

Globalization refers to the long-term trend toward worldwide integration of economic activity and cultures. Data from the World Bank suggest economic openness, as measured by international trade as a percentage of total output, increased steadily from about 25% in the early 1970s to about 60% before the 2008 financial crisis, and has remained near that level since then. We may contrast globalization with **nationalism**—which, in this context, refers to a nation pursuing its own economic interests independently of, or in competition with, the economic interests of other countries.

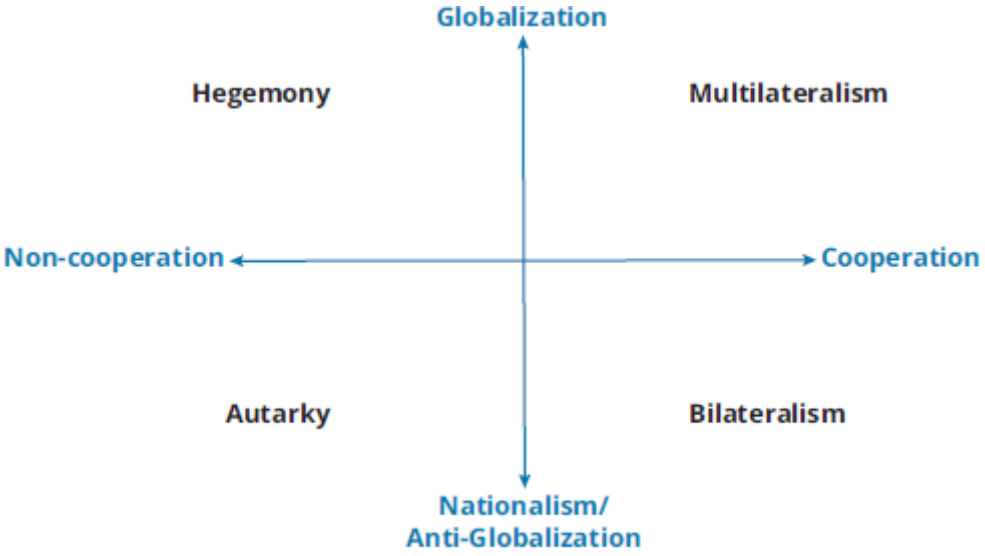


PROFESSOR'S NOTE
Debate about what the word *nationalism* means is beyond the scope of the CFA curriculum. We use it here only in the sense of opposition or resistance to globalization.

As we did with cooperation versus noncooperation, we can think of countries' actions along a spectrum from globalization to nationalism. In general, countries that are closer to the globalization end of the spectrum are those that more actively import and export goods and services, permit freer movement of capital across borders and exchange of currencies, and are more open to cultural interaction.

In Figure 16.1, we draw each spectrum as an axis. This creates four quadrants, each of which we can associate with a type of behavior by countries. While individual countries rarely fit neatly into one of these categories, this gives us a general framework within which we can describe geopolitical actions.

Figure 16.1: Archetypes of Globalization and Cooperation



Source: Reproduced from Level I CFA Curriculum learning module, "Introduction to Geopolitics," with permission from CFA Institute.

Characteristics we may associate with each of these categories are as follows:

- **Autarky** (noncooperation and nationalism) refers to a goal of national self-reliance, including producing most or all necessary goods and services domestically. Autarky is often associated with a state-dominated society in general, with attributes such as government control of industry and media.
- **Hegemony** (noncooperation and globalization) refers to countries that are open to globalization but have the size and scale to influence other countries without necessarily cooperating.
- **Bilateralism** (cooperation and nationalism) refers to cooperation between two countries. A country that engages in bilateralism may have many such relationships with other countries while tending not to involve itself in multicountry arrangements.
- **Multilateralism** (cooperation and globalization) refers to countries that engage extensively in international trade and other forms of cooperation with many other countries. Some countries may exhibit **regionalism**, cooperating multilaterally with nearby countries but less so with the world at large.

Some of the nonstate actors within a country may be more oriented toward globalization than their governments. Businesses may look outside their home country for opportunities to increase profits, reduce costs, and sell to new markets. Investors may seek higher returns or diversification by investing outside their home country. Nonstate actors might buy and sell foreign securities (**portfolio investment flows**) or own physical production capacity in other countries (**foreign direct investment**).

LOS 16.c: Describe functions and objectives of the international organizations that facilitate trade, including the World Bank, the International Monetary Fund,

and the World Trade Organization.

Perhaps the best way to understand the roles of the organizations designed to facilitate trade is to examine their own statements.

The following is according to the **International Monetary Fund (IMF)** (more available at www.IMF.org):

Article I of the Articles of Agreement sets out the IMF's main goals:

- Promoting international monetary cooperation
- Facilitating the expansion and balanced growth of international trade
- Promoting exchange stability
- Assisting in the establishment of a multilateral system of payments
- Making resources available (with adequate safeguards) to members experiencing balance of payments difficulties

The following is according to the **World Bank** (www.worldbank.org):

> The World Bank is a vital source of financial and technical assistance to developing countries around the world. Our mission is to fight poverty with passion and professionalism for lasting results and to help people help themselves and their environment by providing resources, sharing knowledge, building capacity, and forging partnerships in the public and private sectors.
>
> We are not a bank in the common sense; we are made up of two unique development institutions owned by 187 member countries: the International Bank for Reconstruction and Development (IBRD) and the International Development Association (IDA).
>
> Each institution plays a different but collaborative role in advancing the vision of inclusive and sustainable globalization. The IBRD aims to reduce poverty in middle-income and creditworthy poorer countries, while IDA focuses on the world's poorest countries.
>
> […] Together, we provide low-interest loans, interest-free credits, and grants to developing countries for a wide array of purposes that include investments in education, health, public administration, infrastructure, financial and private sector development, agriculture, and environmental and natural resource management.

The following is according to the **World Trade Organization (WTO)** (more available at www.WTO.org):

> The World Trade Organization (WTO) is the only international organization dealing with the global rules of trade between nations. Its main function is to ensure that trade flows as smoothly, predictably, and freely as possible.
>
> […] Trade friction is channeled into the WTO's dispute settlement process where the focus is on interpreting agreements and commitments, and how to ensure that countries' trade policies conform with them. That way, the risk of disputes spilling over into political or military conflict is reduced.
>
> […] At the heart of the system—known as the multilateral trading system—are the WTO's agreements, negotiated and signed by a large majority of the world's trading nations, and ratified in their parliaments. These agreements are the legal ground-rules for international commerce. Essentially, they are contracts, guaranteeing member countries important trade rights. They also bind governments to keep their trade policies within agreed limits to everybody's benefit.

LOS 16.d: Describe geopolitical risk.

Geopolitical risk is the possibility of events that interrupt peaceful international relations. We can classify geopolitical risk into three types:

1. **Event risk** refers to events about which we know the timing but not the outcome, such as national elections.
2. **Exogenous risk** refers to unanticipated events, such as outbreaks of war or rebellion.
3. **Thematic risk** refers to known factors that have effects over long periods, such as human migration patterns or cyber risks.

Geopolitical risk affects investment values by increasing or decreasing the risk premium investors require to hold assets in a country or region. To forecast the effect on investments of a geopolitical risk, we need to consider its probability (*likelihood*), the magnitude of its effects on investment outcomes (*impact*), and how quickly investment values would reflect these effects (*velocity*).

We can use our framework of cooperation and globalization to help estimate the **likelihood of geopolitical risk**. Countries that are more cooperative and globalized tend to have less likelihood of some geopolitical risks, such as armed conflict, but may have greater likelihood of other risks, such as the supply chain disruptions that followed the COVID-19 pandemic in 2020 and 2021.

To analyze the **velocity of geopolitical risk**, we can classify risks as high velocity (short term), medium velocity, or low velocity (long term). Exogenous risks often have high-velocity effects on financial markets and investment values. **Black swan risk** is a term for the risk of low-likelihood exogenous events that have substantial short-term effects. Investors with longer time horizons typically do not need to react to these kinds of events, but investors with shorter horizons might find it necessary to react.

Medium-velocity risks can potentially damage specific companies or industries by increasing their costs or disrupting their production processes, while low-velocity risks tend to affect them in the "environmental, social, and governance" realm. Analyzing these kinds of risk is important for investors with long time horizons.

LOS 16.e: Describe tools of geopolitics and their impact on regions and economies.

We can consider **tools of geopolitics**, the means by which (primarily) state actors advance their interests in the world, as falling into three broad categories of national security, economic, and financial.

National security tools may include armed conflict, espionage, or bilateral or multilateral agreements designed to reinforce or prevent armed conflict. We can say a national security tool is *active* if a country is currently using it or *threatened* if a country is not currently using it but appears likely to do so. Armed conflict affects

regions and economies by destroying productive capital and causing migration away from areas of conflict.

Economic tools can be cooperative or noncooperative. Examples of cooperative economic tools include free trade areas, common markets, and economic and monetary unions (each of which we describe in our reading on international trade and capital flows). Examples of noncooperative economic tools include domestic content requirements, voluntary export restraints, and nationalization (i.e., the state taking control) of companies or industries.

Financial tools include foreign investment and the exchange of currencies. We can view countries as using these tools cooperatively if they allow foreign investment and the free exchange of currencies, or noncooperatively when they restrict these activities. **Sanctions**, or restrictions on a specific geopolitical actor's financial interests, are a financial tool that state actors may use alongside national security tools.

LOS 16.f: Describe the impact of geopolitical risk on investments.

Because analyzing geopolitical risks requires effort, time, and resources, investors should consider whether the **impact of geopolitical risk** is likely to be high or low, and focus their analysis on risks that could have a high impact. With regard to those risks, investors should determine whether they are likely to have *discrete impacts* on a company or industry or *broad impacts* on a country, a region, or the world. Business cycles can affect the impact of geopolitical risk, in that these risks may have greater impacts on investment values when an economy is in recession than they would have during an expansion.

Investors can use qualitative or quantitative **scenario analysis** to gauge the potential effects of geopolitical risks on their portfolios. To help identify geopolitical risks over time, investors may identify **signposts**, or data that can signal when the likelihood of an event is increasing or decreasing, such as volatility indicators in financial markets.



MODULE QUIZ 16.1

1. A state actor that is generally cooperative with other countries and primarily nationalist in pursuing its objectives is *most accurately* said to exhibit:
   A. autarky.
   B. hegemony.
   C. bilateralism.
2. Which of the following tools of geopolitics is *best* described as a noncooperative economic tool?
   A. Voluntary export restraints.
   B. Regional free trade agreements.
   C. Restrictions on conversion of currencies.
3. When investing for a long time horizon, a portfolio manager should *most likely* devote resources to analyzing:
   A. event risks.
   B. thematic risks.

C. exogenous risks.

4. Which international organization is primarily concerned with providing economic assistance to developing countries?
   A. World Bank.
   B. World Trade Organization.
   C. International Monetary Fund.

KEY CONCEPTS

LOS 16.a

Geopolitics refers to interactions among nations. On various issues ranging from diplomacy and military force to economic or cultural openness, countries lie along a spectrum from cooperative to noncooperative.

LOS 16.b

Globalization refers to integration of economic activity and cultures among countries, and can be contrasted with nationalism, which refers to a country pursuing its own interests independently of other countries. Analysts should view geopolitical actions as being on a spectrum from nationalism to globalization.

We may describe geopolitics and its relationship with globalization using the following four broad categories: autarky (noncooperation and nationalism), hegemony (noncooperation and globalization), bilateralism (cooperation and nationalism), and multilateralism (cooperation and globalization).

LOS 16.c

The International Monetary Fund facilitates trade by promoting international monetary cooperation and exchange rate stability, assists in setting up international payments systems, and makes resources available to member countries with balance of payments problems.

The World Bank provides low-interest loans, interest-free credits, and grants to developing countries for many specific purposes. It also provides resources and knowledge and helps form private/public partnerships with the overall goal of fighting poverty.

The World Trade Organization has the goal of ensuring that trade flows freely and works smoothly. Its main focus is on instituting, interpreting, and enforcing numerous multilateral trade agreements that detail global trade policies for a large majority of the world's trading nations.

LOS 16.d

Categories of geopolitical risk are event risk (when the timing is known), exogenous risk (unanticipated events), and thematic risk (known factors that have long-term effects).

LOS 16.e

Tools of geopolitics include national security tools, economic tools, and financial tools.

National security tools may include armed conflict, espionage, or bilateral or multilateral national security agreements.

Cooperative economic tools include free trade areas, common markets, and economic and monetary unions. Noncooperative economic tools include domestic content requirements, voluntary export restraints, and nationalization.

Financial tools include foreign investment, exchange of currencies, and sanctions.

LOS 16.f

Investors should analyze the likelihood of a geopolitical risk, the impact on investment values of an event if it occurs, and the velocity with which it would affect investment values.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 16.1

1. **C** Bilateralism is characterized by nationalism (as contrasted with globalization) and cooperation. Both autarky and hegemony are characterized by noncooperation. (LOS 16.b)

2. **A** Voluntary export restraints (exporting less of a good than the global market demands) are an example of a noncooperative economic tool. Restrictions on the exchange of currencies are a financial tool. Free trade agreements are a cooperative economic tool. (LOS 16.e)

3. **B** Thematic risks are those that have effects over the long term. Event risks and exogenous risks are more likely to have high-velocity impacts on investment values, but they are less of a focus for investors with longer time horizons. (LOS 16.d)

4. **A** The World Bank provides technical and financial assistance to economically developing countries. The World Trade Organization is primarily concerned with settling disputes among countries concerning international trade. The International Monetary Fund promotes international trade and exchange rate stability and assists member countries that experience balance of payments trouble. (LOS 16.c)

READING 17

INTERNATIONAL TRADE

MODULE 17.1: INTERNATIONAL TRADE



Video covering this content is available online.

LOS 17.a: Describe the benefits and costs of international trade.

Historically, economic models of trade have focused on the gains that result when countries with lower relative cost of (a comparative advantage in) the production of a good specialize in producing that good and export it, importing goods for which other countries have a lower relative cost or production. This increases the total output of goods and the wealth of both countries. Comparative advantage results from differences in technology and resource endowments across countries.

Newer models of trade emphasize gains from economies of scale that reduce costs of export goods, an increased variety of goods produced, decreasing costs and improved quality from additional competition, and more efficient allocation of productive resources.

Free trade can also benefit consumers by reducing the pricing power of domestic monopolies. Countries that produce the same good can also operate in market of monopolistic competition with differentiated products. Consider the global market for automobiles. Several countries export one type of autos and import other types. Consumers benefit from the greater variety offered, as well as the reduced costs from specialization.

While domestic consumers of imported goods and domestic producers of exported goods both gain, international trade imposes costs as well. The most cited costs of free trade are the loss of domestic jobs in an importing industry and increased economic inequality. Consider a country with a comparative advantage in the production of steel that exports to another country with a higher relative cost of steel production. When free trade is permitted, consumers of steel in the importing country gain and workers and producers of steel lose. In the exporting country, increased demand for steel production may increase prices to domestic consumers, while workers and companies in the steel industry will gain, at least in the short run.

In a situation where labor costs are lower in the exporting country, free trade can decrease wages and employment in the domestic economy. Consider textile production in a country with a relatively high cost of labor. For a labor-intensive industry such as some textile production, importing textiles from a lower-cost country will result in job and wage losses in the domestic industry, possibly increasing income inequality.

When we speak of the gains from trade in general, it is based on economic analysis suggesting that, overall, the gains from trade are greater than the losses—especially in the long run. That is, the gainers could, at least in theory, compensate the losers, with net gains shared by both importing and exporting countries. While free trade surely imposes costs on the workers and companies in an industry facing competition from imports in the short run, it is argued that in the long run—after workers receive additional training and find work in other industries—the short-run costs of free trade are mitigated to some significant degree, or even reversed.

LOS 17.b: Compare types of trade restrictions, such as tariffs, quotas, and export subsidies, and their economic implications.

There are many reasons (at least stated reasons) given by governments that impose trade restrictions. Some have support among economists as conceivably valid in terms of increasing a country's welfare, while others get little or no support from economic theory. Some of the reasons for trade restrictions that have support from economists are as follows:

- *Infant industry.* Protection from foreign competition is given to new industries to give them an opportunity to grow to an internationally competitive scale and get up the learning curve in terms of efficient production methods.
- *National security.* Even if imports are cheaper, it may be in the country's best interest to protect producers of goods crucial to the country's national defense so that those goods are available domestically in the event of conflict.

Other arguments for trade restrictions that have little support in theory are as follows:

- *Protecting domestic jobs.* While some jobs are certainly lost, and some groups and regions are negatively affected by free trade, other jobs (in export industries or growing domestic goods and services industries) will be created, and prices for domestic consumers will be less without import restrictions.
- *Protecting domestic industries*. Industry firms often use political influence to get protection from foreign competition—usually to the detriment of consumers, who pay higher prices.

Other arguments include retaliation for foreign trade restrictions, government collection of tariffs (like taxes on imported goods), countering the effects of government subsidies paid to foreign producers, and preventing foreign exports at less than their cost of production (*dumping*).

Types of trade restrictions include the following:

- **Tariffs.** These are taxes on imported goods collected by the government.
- **Quotas.** These are limits on the amount of imports allowed over some period.
- **Export subsidies.** These are government payments to firms that export goods.
- **Minimum domestic content.** This is the requirement that some percentage of product content must be from the domestic country.

- **Voluntary export restraint.** A country voluntarily restricts the amount of a good that can be exported, often in the hope of avoiding tariffs or quotas imposed by its trading partners.

Economic Implications of Trade Restrictions

We will now examine the effects of the primary types of trade restrictions, tariffs, and subsidies.

A **tariff** placed on an imported good increases its domestic price, decreases the quantity imported, and increases the quantity supplied domestically. Domestic producers gain, foreign exporters lose, and the domestic government gains by the amount of the tariff revenues.

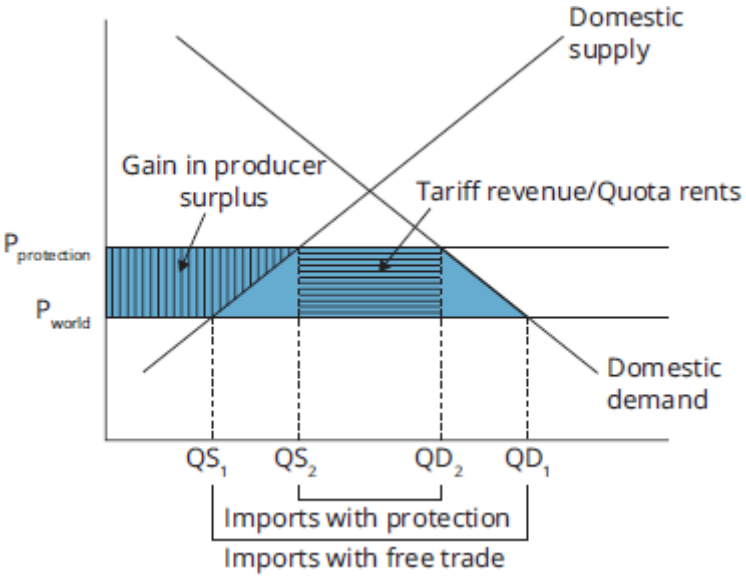
A **quota** restricts the quantity of a good imported to the quota amount. Domestic producers gain, and domestic consumers lose from an increase in the domestic price. The right to export a specific quantity to the domestic country is granted by the domestic government, which may or may not charge for the import licenses to foreign countries. If the import licenses are sold, the domestic government gains the revenue.

We illustrate the overall welfare effects of quotas and tariffs for a small country in Figure 17.1. We define a quota that is equivalent to a given tariff as a quota that will result in the same decrease in the quantity of a good imported as the tariff. Defined this way, a tariff and an equivalent quota both increase the domestic price from P_{world}, the price that prevails with no trade restriction, to $P_{\text{protection}}$.

At P_{world}, before any restriction, the domestic quantity supplied is QS_1, and the domestic quantity demanded is QD_1, with the difference equal to the quantity imported, $\text{QD}_1 - \text{QS}_1$. Placing a tariff on imports increases the domestic price to $P_{\text{protection}}$, increases the domestic quantity supplied to QS_2, and decreases the domestic quantity demanded to QD_2. The difference is the new quantity imported. An equivalent quota will have the same effect, decreasing the quantity imported to $\text{QD}_2 - \text{QS}_2$.

The entire shaded area in Figure 17.1 represents the loss of consumer surplus in the domestic economy. The portion with vertical lines, the area to the left of the domestic supply curve between $P_{\text{protection}}$ and P_{world}, represents the gain in the producer surplus of domestic producers. The portion with horizontal lines, the area bounded by $\text{QD}_2 - \text{QS}_2$ and $\text{P}_{\text{protection}} - \text{P}_{\text{world}}$, represents the gain to the domestic government from tariff revenue. The two remaining triangular areas are the deadweight loss from the restriction on free trade.

Figure 17.1: Effects of Tariffs and Quotas



In the case of a quota, if the domestic government collects the full value of the import licenses, the result is the same as for a tariff. If the domestic government does not charge for the import licenses, this amount is a gain to those foreign exporters who receive the import licenses under the quota and are termed **quota rents**.

In terms of overall economic gains from trade, the deadweight loss is the amount of lost welfare from the imposition of a quota or tariff. From the viewpoint of the domestic country, the loss in consumer surplus is only partially offset by the gains in domestic producer surplus and the collection of tariff revenue.

If none of the quota rents are captured by the domestic government, the overall welfare loss to the domestic economy is greater by the amount of the quota rents. It is the entire difference between the gain in producer surplus and the loss of consumer surplus.

A **voluntary export restraint (VER)** is just as it sounds. It refers to a voluntary agreement by a government to limit the quantity of a good that can be exported. VERs are another way of protecting the domestic producers in the importing country. They result in a welfare loss to the importing country equal to that of an equivalent quota with no government charge for the import licenses—that is, no capture of the quota rents.

Export subsidies are payments by a government to its country's exporters. Export subsidies benefit producers (exporters) of the good but increase prices and reduce consumer surplus in the exporting country. In a small country, the price will increase by the amount of the subsidy to equal the world price plus the subsidy. In the case of a large exporter of the good, the world price decreases and some benefits from the subsidy accrue to foreign consumers, while foreign producers are negatively affected.

Most of the effects of all four of these protectionist policies are the same. With respect to the domestic (importing) country, import quotas, tariffs, and VERs all do the following:

- Reduce imports
- Increase price
- Decrease consumer surplus
- Increase domestic quantity supplied
- Increase producer surplus

With one exception, all will decrease national welfare. Quotas and tariffs in a large country could increase national welfare under a specific set of assumptions, primarily because for a country that imports a large amount of the good, setting a quota or tariff could reduce the world price for the good.

Capital Restrictions

Some countries impose **capital restrictions** on the flow of financial capital across borders. Restrictions include outright prohibition of investment in the domestic country by foreigners, prohibition of or taxes on the income earned on foreign investments by domestic citizens, prohibition of foreign investments in certain domestic industries, and restrictions on repatriation of earnings of foreign entities operating in a country.

Overall, capital restrictions are thought to decrease economic welfare. However, over the short term, they have helped developing countries avoid the impact of great inflows of foreign capital during periods of optimistic expansion and the impact of large outflows of foreign capital during periods of correction and market unease, or outright panic. Even these short-term benefits may not offset longer-term costs if the country is excluded from international markets for financial capital flows.

LOS 17.c: Explain motivations for and advantages of trading blocs, common markets, and economic unions.

There are various types of agreements among countries with respect to trade policy. The essence of all of them is to reduce trade barriers among the countries. Reductions in trade restrictions among countries have some, by now familiar, positive and negative effects on economic welfare. The positive effects result from increased trade according to comparative advantage, as well as increased competition among firms in member countries. The negative effects result because some firms, some industries, and some groups of workers will see their wealth and incomes decrease. Workers in affected industries may need to learn new skills to get new jobs.

On balance, economic welfare is improved by reducing or eliminating trade restrictions. Note, however, that to the extent that a trade agreement increases trade restrictions on imports from nonmember countries, economic welfare gains are reduced and, in an extreme case, could be outweighed by the costs such restrictions impose. This could result if restrictions on trade with nonmember countries increase a country's (unrestricted) imports from a member that has higher prices than the country's previous imports from a nonmember.

We list these types of agreements, generally referred to as **trading blocs** or **regional trading agreements (RTAs)**, in order of their degrees of integration.

Free Trade Areas

1. All barriers to import and export of goods and services among member countries are removed.

Customs Union

1. All barriers to import and export of goods and services among member countries are removed.
2. All countries adopt a common set of trade restrictions with nonmembers.

Common Market

1. All barriers to import and export of goods and services among the countries are removed.
2. All countries adopt a common set of trade restrictions with nonmembers.
3. All barriers to the movement of labor and capital goods among member countries are removed.

Economic Union

1. All barriers to import and export of goods and services among the countries are removed.
2. All countries adopt a common set of trade restrictions with nonmembers.
3. All barriers to the movement of labor and capital goods among member countries are removed.
4. Member countries establish common institutions and economic policy for the union.

Monetary Union

1. All barriers to import and export of goods and services among the countries are removed.
2. All countries adopt a common set of trade restrictions with nonmembers.
3. All barriers to the movement of labor and capital goods among member countries are removed.
4. Member countries establish common institutions and economic policy for the union.
5. Member countries adopt a single currency.

The North American Free Trade Agreement (NAFTA) is an example of a free trade area, the European Union (EU) is an example of an economic union, and the Eurozone is an example of a monetary union.



MODULE QUIZ 17.1

1. Which of the following effects is *most likely* to occur in a country that increases its openness to international trade?
   A. Increased prices of consumer goods.
   B. Greater specialization in domestic output.

C. Decreased employment in exporting industries.

2. An agreement with another country to limit the volume of goods and services sold to them is *best* described as a:
   A. quota.
   B. voluntary export restraint.
   C. minimum domestic content rule.
3. Which of the following groups would be *most likely* to suffer losses from the imposition of a tariff on steel imports?
   A. Domestic steel producers.
   B. Workers in the domestic auto industry.
   C. Workers in the domestic steel industry.
4. The *most likely* motivation for establishing a trading bloc is to:
   A. increase economic welfare in the member countries.
   B. increase tariff revenue for the member governments.
   C. protect domestic industries in the member economies.
5. In which type of regional trade agreement are economic policies conducted independently by the member countries, while labor and capital are free to move among member countries?
   A. Free trade area.
   B. Common market.
   C. Economic union.

KEY CONCEPTS

LOS 17.a

Free trade among countries increases overall economic welfare. Countries can benefit from trade because one country can specialize in the production of an export good and benefit from economies of scale. Economic welfare can also be increased by greater product variety, more competition, and a more efficient allocation of resources.

Costs of free trade are primarily losses to those in domestic industries that lose business to foreign competition, especially less efficient producers who leave an industry. While other domestic industries will benefit from freer trade policies, unemployment may increase over the period in which workers are retrained for jobs in the expanding industries. Some argue that greater income inequality may result, but overall, the gains from liberalization of trade policies are thought to exceed the costs, so that the winners could conceivably compensate the losers and still be better off.

LOS 17.b

Types of trade restrictions include the following:

- *Tariffs*. These are taxes on imported goods collected by the government.
- *Quotas*. These are limits on the amount of imports allowed over some period.
- *Minimum domestic content*. This is the requirement that some percentage of product content must be from the domestic country.
- *Voluntary export restraints*. A country voluntarily restricts the amount of a good that can be exported, often in the hope of avoiding tariffs or quotas imposed by their trading partners.

Within each importing country, all of these restrictions will tend to do the following:

- Increase prices of imports and decrease quantities of imports
- Increase demand for and quantity supplied of domestically produced goods
- Increase producer's surplus and decrease consumer surplus

Export subsidies decrease export prices and benefit importing countries at the expense of the government of the exporting country.

Restrictions on the flow of financial capital across borders include outright prohibition of investment in the domestic country by foreigners, prohibition of or taxes on the income earned on foreign investments by domestic citizens, prohibition of foreign investments in certain domestic industries, and restrictions on repatriation of earnings of foreign entities operating in a country.

LOS 17.c

Trade agreements, which increase economic welfare by facilitating trade among member countries, take the following forms:

- *Free trade area*. All barriers to the import and export of goods and services among member countries are removed.
- *Customs union*. Member countries *also* adopt a common set of trade restrictions with nonmembers.
- *Common market*. Member countries *also* remove all barriers to the movement of labor and capital goods among members.
- *Economic union*. Member countries *also* establish common institutions and economic policy for the union.
- *Monetary union*. Member countries *also* adopt a single currency to reduce the cost of cross-border trade.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 17.1

1. **B** Openness to international trade increases specialization as production shifts to those products in which domestic producers have a comparative advantage. Greater competition from imports will tend to decrease prices for consumer goods. Increasing international trade is likely to increase profitability and employment in exporting industries, but it may decrease profitability and employment in industries that compete with imported goods. (LOS 17.a)

2. **B** Voluntary export restraints are agreements to limit the volume of goods and services exported to another country. Minimum domestic content rules are limitations imposed by a government on its domestic firms. Import quotas are limitations on imports, not on exports. (LOS 17.b)

3. **B** Imposing a tariff on steel imports benefits domestic steel producers and workers by increasing the domestic price of steel and benefits the national government by

increasing tax (tariff) revenue. However, the increase in the domestic price of steel would increase costs in industries that use significant amounts of steel, such as the automobile industry. The resulting increase in the price of automobiles reduces the quantity of automobiles demanded, and it ultimately reduces employment in that industry. (LOS 17.b)

4. **A** The motivation for trading blocs is to increase economic welfare in the member countries by eliminating barriers to trade. Joining a trading bloc may have negative consequences for some domestic industries and may decrease tariff revenue for the government. (LOS 17.c)

5. **B** These characteristics describe a common market. In a free trade area, member countries remove restrictions on goods and services and trade with one another but may still restrict movement of labor and capital among member countries. In an economic union, member countries also coordinate their economic policies and institutions. (LOS 17.c)

READING 18

CAPITAL FLOWS AND THE FX MARKET

MODULE 18.1: THE FOREIGN EXCHANGE MARKET



Video covering this content is available online.

LOS 18.a: Describe the foreign exchange market, including its functions and participants, distinguish between nominal and real exchange rates, and calculate and interpret the percentage change in a currency relative to another currency.

Foreign currency markets serve companies and individuals that purchase or sell foreign goods and services denominated in foreign currencies. An even larger market, however, exists for capital flows. Foreign currencies are needed to purchase foreign physical assets as well as foreign financial securities.

Many companies have foreign exchange risk arising from their cross-border transactions. A Japanese company that expects to receive 10 million euros when a transaction is completed in 90 days has yen/euro exchange rate risk as a result. By entering into a **forward currency contract** to sell 10 million euros in 90 days for a specific quantity of yen, the firm can reduce or eliminate the foreign exchange risk associated with the transaction. When a firm takes a position in the foreign exchange market to reduce an existing risk, we say the firm is **hedging** its risk.

Alternatively, when a transaction in the foreign exchange markets increases currency risk, we describe the transaction or position as **speculating**. Investors, companies, and financial institutions, such as banks and investment funds, all regularly enter into speculative foreign currency transactions.

The primary dealers in foreign exchange (FX) and originators of forward FX contracts are large multinational banks. This part of the FX market is often called the **sell side**. On the other hand, the **buy side** consists of the many buyers of foreign currencies and forward FX contracts. These buyers include the following:

- *Corporations* regularly engage in cross-border transactions, purchase and sell foreign currencies as a result, and enter into FX forward contracts to hedge the risk of expected future receipts and payments denominated in foreign currencies.

- *Investment accounts* of many types transact in foreign currencies, hold foreign securities, and may both speculate and hedge with currency derivatives. **Real money accounts** refer to mutual funds, pension funds, insurance companies, and other institutional accounts that do not use derivatives. **Leveraged accounts** refer to the various types of investment firms that use derivatives, including hedge funds, firms that trade for their own accounts, and other trading firms of various types.
- *Governments and government entities*, including sovereign wealth funds and pension funds, acquire foreign exchange for transactional needs, investment, or speculation. Central banks sometimes engage in FX transactions to affect exchange rates in the short term in accordance with government policy.
- The **retail FX market** refers to FX transactions by households and relatively small institutions, and it may be for tourism, cross-border investment, or speculative trading.

Types of Exchange Rates

An **exchange rate** is simply the price or cost of units of one currency in terms of another. In this book we state exchange rates in the form 1.416 USD/EUR, to mean that each euro costs $1.416. If you read the *"/"* as *per*, you will have no trouble with the notation. We say the exchange rate is $1.416 per euro.



PROFESSOR'S NOTE
This is the way exchange rates are quoted in the Level I CFA curriculum. Foreign exchange traders typically use the inverse (an indirect quotation).

In a foreign currency quotation, we have the price of one currency in units of another currency. These are often referred to as the **base currency** and the **price currency**. In the quotation 1.25 USD/EUR, the USD is the price currency and the EUR is the base currency. The price of one euro (base currency) is 1.25 USD (the price currency), so 1.25 is the price of one unit of the base currency in terms of the other. It may help to remember that the euro, in this example, is in the bottom or "base" of the exchange rate given in terms of USD/EUR.

An exchange rate expressed as price currency/base currency is referred to as a **direct quote** from the point of view of an investor in the price currency country and an **indirect quote** from the point of view of an investor in the base currency country. For example, a quote of 1.17 USD/EUR would be a direct quote for a USD-based investor and an indirect quote for a EUR-based investor. Conversely, a quote of 1 / 1.17 = 0.855 EUR/USD would be a direct quote for a EUR-based investor and an indirect quote for a USD-based investor.

The exchange rate at a point in time is referred to as a **nominal exchange rate**. If this rate (price/base) increases, the cost of a unit of the base currency in terms of the price currency has increased, so that the purchasing power of the price currency has decreased. If the USD/EUR exchange rate increases from 1.10 to 1.15, the cost of 100 euros increases from $110 to $115. The purchasing power of the dollar has decreased relative to the euro because the cost of 100 euros' worth of goods to a consumer in the United States has increased over the period.

The purchasing power of one currency relative to another is also affected by changes in the price levels of the two countries. The **real exchange rate** between two currencies refers to the purchasing power of one currency in terms of the amount of goods priced in another currency, relative to an earlier (base) period.

Consider a situation in which the nominal USD/EUR exchange rate is unchanged at 1.00 over a period and the price level in the United States is unchanged, while prices in the Eurozone have increased by 5%. Eurozone goods that cost 100 euros at the beginning of the period cost 105 euros at the end of the period. With the nominal exchange rate unchanged, the purchasing power of the USD in the Eurozone has decreased, because exchanging 100 USD for 100 EUR will now buy only 100 / 105 = 95.2% of the goods 100 EUR could buy at the beginning of the period.

Here is a summary:

- An increase in the *nominal* USD/EUR rate decreases the purchasing power of the USD in the Eurozone (and increases the purchasing power of the EUR in the United States); the *real* USD/EUR exchange rate has increased.
- A decrease in the *nominal* USD/EUR rate increases the purchasing power of the USD in the Eurozone (and decreases the purchasing power of the EUR in the United States); the *real* USD/EUR exchange rate has decreased.
- An increase in the Eurozone price level, relative to the price level in the United States, will increase the *real* USD/EUR exchange rate, decreasing the purchasing power of the USD in the Eurozone (and increasing the purchasing power of the EUR in the United States).
- A decrease in the Eurozone price level, relative to the price level in the United States, will decrease the *real* USD/EUR exchange rate, increasing the purchasing power of the USD in the Eurozone (and decreasing the purchasing power of the EUR in the United States).

The end-of-period real P/B exchange rate can be calculated as follows:

$$\text{real P/B exchange rate} = \text{nominal P/B exchange rate} \times \left(\frac{\text{CPI}_{\text{base currency}}}{\text{CPI}_{\text{price currency}}} \right)$$

where the CPI values are relative to base period values of 100.

We can see from the formula the following:

- An increase (decrease) in the nominal exchange rate over the period increases (decreases) the end-of period real exchange rate, and the purchasing power of the price currency decreases (increases).
- An increase in the price level in the price currency country relative to the price level in the base currency country will decrease the real exchange rate, increasing the purchasing power of the price currency in terms of base country goods.
- Conversely, a decrease in the price level in the price currency country relative to the price level in the base currency country will increase the real exchange rate, decreasing the purchasing power of the price currency in terms of base country goods.

In the following example, we calculate the end-of-period real $/£ exchange rate when the nominal $/£ exchange rate has decreased over the period (which tends to decrease the real exchange rate and increase the purchasing power of the price currency)—and when the price level in the United Kingdom has increased by more than the price level in the United States over the period (which tends to increase the real exchange rate and decrease the purchasing power of the price currency). The relative increase in U.K. prices has reduced the effects of the decrease in the nominal exchange rate on the increase in the purchasing power of the USD.

> **EXAMPLE: Real exchange rate**
>
> At a base period, the CPIs of the United States and United Kingdom are both 100, and the exchange rate is $1.70/£. Three years later, the exchange rate is $1.60/£, and the CPI has risen to 110 in the United States and 112 in the United Kingdom. What is the real exchange rate at the end of the three-year period?
>
> **Answer:**
>
> The real exchange rate is $1.60 / £ × 112 / 110 = $1.629 / £, which means that U.S. goods and services that cost $1.70 at the base period now cost only $1.629 (in real terms) if purchased in the United Kingdom and the real exchange rate, $/£, has fallen. The decrease in the real exchange rate (and the increase in the purchasing power of the USD in terms of U.K. goods) over the period is less than it would have been if the relative prices between the two countries had not changed.

A **spot exchange rate** is the currency exchange rate for immediate delivery, which for most currencies means the exchange of currencies takes place two days after the trade.

A **forward exchange rate** is a currency exchange rate for an exchange to be done in the future. Forward rates are quoted for various future dates (e.g., 30 days, 60 days, 90 days, or one year). A forward is actually an agreement to exchange a specific amount of one currency for a specific amount of another on a future date specified in the forward agreement.

A French firm that will receive 10 million GBP from a British firm six months from now has uncertainty about the amount of euros that payment will be equivalent to six months from now. By entering into a forward agreement covering 10 million GBP at the 6-month forward rate of 1.192 EUR/GBP, the French firm has agreed to exchange 10 million GBP for 11.92 million euros in six months.

Calculating the Percentage Change in the FX Value of a Currency

Consider a USD/EUR exchange rate that has changed from 1.42 to 1.39 USD/EUR. The percentage change in the dollar price of a euro is simply 1.39 / 1.42 − 1 = −0.0211 = −2.11%. Because the dollar price of a euro has fallen, the euro has *depreciated* relative to the dollar, and a euro now buys 2.11% fewer U.S. dollars. It is correct to say that the euro has depreciated by 2.11% relative to the dollar.

On the other hand, it is *not* correct to say that the dollar has appreciated by 2.11%. To calculate the percentage appreciation of the dollar, we need to convert the quotes to EUR/USD. So, our beginning quote of 1.42 USD/EUR becomes 1 / 1.42 = 0.7042 EUR/USD, and our ending quote of 1.39 USD/EUR becomes 1 / 1.39 = 0.7194 EUR/USD. Using these exchange rates, we can calculate the change in the euro price of a dollar as 0.7194 / 0.7042 − 1 = 0.0216 = 2.16%. In this case, it is correct to say that the dollar has appreciated 2.16% with respect to the euro. Note that for the given exchange rate quotes, the percentage appreciation of the dollar is not the same as the percentage depreciation in the euro.

MODULE QUIZ 18.1

1. One year ago, the nominal exchange rate for USD/EUR was 1.300. Since then, the real exchange rate has increased by 3%. This *most likely* implies that:
   A. the nominal exchange rate is less than USD/EUR 1.235.
   B. the purchasing power of the euro has increased approximately 3% in terms of U.S. goods.
   C. inflation in the euro zone was approximately 3% higher than inflation in the United States.
2. Sell-side participants in the foreign exchange market are *most likely* to include:
   A. banks.
   B. hedge funds.
   C. insurance companies.
3. Suppose that the quote for British pounds (GBP) in New York is USD/GBP 1.3110. What is the quote for U.S. dollars (USD) in London (GBP/USD)?
   A. 0.3110.
   B. 0.7628.
   C. 1.3110.
4. The Canadian dollar (CAD) exchange rate with the Japanese yen (JPY) changes from JPY/CAD 75 to JPY/CAD 78. The CAD has:
   A. depreciated by 3.8%, and the JPY has appreciated by 4.0%.
   B. appreciated by 3.8%, and the JPY has depreciated by 4.0%.
   C. appreciated by 4.0%, and the JPY has depreciated by 3.8%.

MODULE 18.2: MANAGING EXCHANGE RATES

Video covering this content is available online.

LOS 18.b: Describe exchange rate regimes and explain the effects of exchange rates on countries' international trade and capital flows.

The IMF categorizes **exchange rate regimes** into the following types, which include two for countries that do not issue their own currencies and seven for countries that issue their own currencies.

Countries That Do Not Have Their Own Currency

A country can use the currency of another country (**formal dollarization**). The country cannot have its own monetary policy, as it does not create money or issue currency.

A country can be a member of a **monetary union** in which several countries use a common currency. Within the European Union, for example, most countries use the euro. While individual countries give up the ability to set domestic monetary policy, they all participate in determining the monetary policy of the European Central Bank.

Countries That Have Their Own Currency

A **currency board arrangement** is an explicit commitment to exchange domestic currency for a specified foreign currency at a fixed exchange rate. A notable example of such an arrangement is Hong Kong. In Hong Kong, currency is (and may be) only issued when fully backed by holdings of an equivalent amount of U.S. dollars. The Hong Kong Monetary Authority can earn interest on its U.S. dollar balances. With dollarization, there is no such income, as the income is earned by the U.S. Federal Reserve when it buys interest-bearing assets with the U.S. currency it issues. While the monetary authority gives up the ability to conduct independent monetary policy and essentially imports the inflation rate of the outside currency, there may be some latitude to affect interest rates over the short term.

In a **conventional fixed peg arrangement**, a country pegs its currency within margins of ±1% versus another currency or a basket that includes the currencies of its major trading or financial partners. The monetary authority can maintain exchange rates within the band by purchasing or selling foreign currencies in the foreign exchange markets (*direct intervention*). In addition, the country can use *indirect intervention*, including changes in interest rate policy, regulation of foreign exchange transactions, and convincing people to constrain foreign exchange activity. The monetary authority retains more flexibility to conduct monetary policy than with dollarization, a monetary union, or a currency board. However, changes in policy are constrained by the requirements of the peg.

In a system of pegged exchange rates within horizontal bands or a **target zone**, the permitted fluctuations in currency value relative to another currency or basket of currencies are wider (e.g., ±2%). Compared to a conventional peg, the monetary authority has more policy discretion because the bands are wider.

With a **crawling peg**, the exchange rate is adjusted periodically, typically to adjust for higher inflation versus the currency used in the peg. This is termed a *passive crawling peg* as opposed to an *active crawling peg,* in which a series of exchange rate adjustments over time is announced and implemented. An active crawling peg can influence inflation expectations, adding some predictability to domestic inflation. Monetary policy is restricted in much the same way it is with a fixed peg arrangement.

With **management of exchange rates within crawling bands**, the width of the bands that identify permissible exchange rates is increased over time. This method can be used to transition from a fixed peg to a floating rate when the monetary authority's lack of credibility makes an immediate change to floating rates impractical. Again, the degree of monetary policy flexibility increases with the width of the bands.

With a system of **managed floating exchange rates**, the monetary authority attempts to influence the exchange rate in response to specific indicators such as the balance of payments, inflation rates, or employment without any specific target exchange rate or

predetermined exchange rate path. Intervention may be direct or indirect. Such management of exchange rates may induce trading partners to respond in ways that reduce stability.

When a currency is **independently floating**, the exchange rate is market determined, and foreign exchange market intervention is used only to slow the rate of change and reduce short-term fluctuations—not to keep exchange rates at a target level.

Changes in Exchange Rates

Changes in exchange rates can affect both imports and exports. If the USD/EUR exchange rate decreases, the USD has appreciated relative to the EUR. In this case, the USD prices of imports from the Eurozone have decreased, and it takes more euros to purchase a U.S. good at the new exchange rate. As a result, Eurozone imports from the United States (U.S. exports to the Eurozone) will decrease, and U.S. imports of Eurozone goods (Eurozone exports to the United States) will increase.

The effects on the goods market (imports and exports of merchandise) occur more slowly than the effects on capital flows between countries. This is reflected in a country's **balance of payments**. Essentially, the balance of payments refers to the fact that capital flows must offset any imbalance between the value of a country's exports to and imports from another country.

Consider two countries that have an imbalance between the value of their exports to and imports from another country, such as between China and the United States. Typically, U.S. imports from China have a greater value than U.S. exports to China. We say that the United States has a goods and services **trade deficit** with China and China has a trade surplus with the United States. Capital flows must offset this difference.

Capital flows result primarily from purchases of assets (both physical and financial) in one country by investors or governments in other countries. A large portion of the capital flows between the United States and China are purchases of U.S. debt securities by China. China has a **capital account deficit** that offsets its trade surplus with the United States.

The relation between the balance of trade and capital flows is expressed by the following identity, which was presented in the Economics prerequisite readings:

$$\begin{aligned}&(\text{exports} - \text{imports}) \equiv (\text{private savings} - \text{investment in physical capital})\\&+ (\text{tax revenue} - \text{government spending})\end{aligned}$$

or:

$$(X - M) \equiv (S - I) + (T - G)$$

The intuition is that a trade deficit ($X - M < 0$) means that the right-hand side must also be negative, which implies that total domestic savings (private savings + government savings) are less than domestic investment in physical capital. The additional amount to fund domestic investment must come from foreigners, which results in a surplus in the capital account to offset the deficit in the trade account. Another thing we can see from this identity is that any government deficit not funded by an excess of domestic

saving over domestic investment is consistent with a trade deficit (imports > exports), which is offset by an inflow of foreign capital (a surplus in the capital account).

Capital flows adjust more rapidly than spending, savings, and asset prices, so they are the primary determinant of exchange rates in the short and intermediate term. Trade flows are more important in determining exchange rates in the long term as asset prices and saving/investment decisions adjust over time.

LOS 18.c: Describe common objectives of capital restrictions imposed by governments.

Governments sometimes place restrictions on the flow of investment capital into their country, out of their country, or both. Commonly cited objectives of capital flow restrictions include the following:

- *Reduce the volatility of domestic asset prices.* In times of macroeconomic crisis, capital flows out of the country can drive down asset prices drastically, especially with prices of liquid assets such as stocks and bonds. With no restrictions on inflows or outflows of foreign investment capital, the asset markets of countries with economies that are small relative to the amount of foreign investment can be quite volatile over a country's economic cycle.
- *Maintain fixed exchange rates.* For countries with fixed exchange rate targets, limiting flows of foreign investment capital makes it easier to meet the exchange rate target—and, therefore, to be able to use monetary and fiscal policy to pursue only the economic goals for the domestic economy.
- *Keep domestic interest rates low.* By restricting the outflow of investment capital, countries can keep their domestic interest rates low and manage the domestic economy with monetary policy, as investors cannot pursue higher rates in foreign countries. China is an example of a country with a fixed exchange rate regime where restrictions on capital flows allow policymakers to maintain the target exchange rate as well as to pursue a monetary policy independent of concerns about its effect on currency exchange rates.
- *Protect strategic industries.* Governments sometimes prohibit investment by foreign entities in industries considered to be important for national security, such as the telecommunications and defense industries.



MODULE QUIZ 18.2

1. The monetary authority of the Stoddard Islands will exchange its currency for U.S. dollars at a one-for-one ratio. As a result, the exchange rate of the Stoddard Islands currency with the U.S. dollar is 1.00, and many businesses in the Islands will accept U.S. dollars in transactions. This exchange rate regime is *best* described as:
   A. a fixed peg.
   B. dollarization.
   C. a currency board.

2. Other things equal, which of the following is *most likely* to decrease a country's trade deficit?
   A. Increase the capital account surplus.
   B. Decrease expenditures relative to income.

C. Decrease domestic saving relative to domestic investment.

3. The goal of a government that imposes restrictions on foreign capital flows is *most likely* to:
   A. stimulate domestic interest rates.
   B. decrease domestic asset price volatility.
   C. encourage competition with domestic industries.

KEY CONCEPTS

LOS 18.a

Currency exchange rates are given as the price of one unit of currency in terms of another. A nominal exchange rate of 1.44 USD/EUR is interpreted as $1.44 per euro. We refer to the USD as the price currency and the EUR as the base currency.

An increase (decrease) in an exchange rate represents an appreciation (depreciation) of the base currency relative to the price currency.

A spot exchange rate is the rate for immediate delivery. A forward exchange rate is a rate for exchange of currencies at some future date.

A real exchange rate measures changes in relative purchasing power over time:

$$\text{real exchange rate} = \text{nominal exchange rate} \times \left(\frac{\text{CPI}_{\text{base currency}}}{\text{CPI}_{\text{price currency}}} \right)$$

For a change in an exchange rate, we can calculate the percentage appreciation (price goes up) or depreciation (price goes down) of the base currency. For example, a decrease in the USD/EUR exchange rate from 1.44 to 1.42 represents a depreciation of the EUR relative to the USD of 1.39% (1.42 / 1.44 − 1 = −0.0139) because the value of a euro has fallen 1.39%.

Given the depreciation of the EUR (base currency) of 1.39%, we can calculate the appreciation of the price currency as $\frac{1}{(1 - 0.0139)} - 1 = 0.0141 = 1.41\%$.

The market for foreign exchange is the largest financial market in terms of the value of daily transactions and has various participants, including large multinational banks (the sell side) and corporations, investment fund managers, hedge fund managers, investors, governments, and central banks (the buy side).

Participants in the foreign exchange markets are referred to as hedgers if they enter into transactions that decrease an existing foreign exchange risk, and as speculators if they enter into transactions that increase their foreign exchange risk.

LOS 18.b

Exchange rate regimes for countries that do not have their own currency:

- With *formal dollarization*, a country uses the currency of another country.
- In a *monetary union*, several countries use a common currency.

Exchange rate regimes for countries that have their own currency:

- A *currency board arrangement* is an explicit commitment to exchange domestic currency for a specified foreign currency at a fixed exchange rate.
- In a *conventional fixed peg arrangement*, a country pegs its currency within margins of ±1% versus another currency.
- In a system of *pegged exchange rates within horizontal bands* or a *target zone*, the permitted fluctuations in currency value relative to another currency or basket of currencies are wider (e.g., ±2%).
- With a *crawling peg*, the exchange rate is adjusted periodically, typically to adjust for higher inflation versus the currency used in the peg.
- With *management of exchange rates within crawling bands*, the width of the bands that identify permissible exchange rates is increased over time.
- With a system of *managed floating exchange rates*, the monetary authority attempts to influence the exchange rate in response to specific indicators, such as the balance of payments, inflation rates, or employment, without any specific target exchange rate.
- When a currency is *independently floating*, the exchange rate is market determined.

LOS 18.c

Commonly cited objectives of capital flow restrictions include the following:

- Reducing the volatility of domestic asset prices
- Maintaining fixed exchange rates
- Keeping domestic interest rates low and enabling greater independence regarding monetary policy
- Protecting strategic industries from foreign ownership

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 18.1

1. **B** An increase in the real exchange rate USD/EUR (the number of USD per one EUR) means that a euro is worth more in purchasing power (real) terms in the United States. Changes in a real exchange rate depend on the change in the nominal exchange rate relative to the difference in inflation. By itself, a real exchange rate does not indicate the directions or degrees of change in either the nominal exchange rate or the inflation difference. (LOS 18.a)

2. **A** Large multinational banks make up the sell side of the foreign exchange market. The buy side includes corporations, real money and leveraged investment accounts, governments and government entities, and retail purchasers of foreign currencies. (LOS 18.a)

3. **B** 1 / 1.311 = 0.7628 GBP/USD

   (LOS 18.a)

4. **C** The CAD has appreciated because it is worth a larger number of JPY. The percentage appreciation is (78 − 75) / 75 = 4.0%. To calculate the percentage depreciation of the JPY against the CAD, convert the exchange rates to make JPY the base currency: 1 / 75 = 0.0133 CAD/JPY and 1 / 78 = 0.0128 CAD/JPY. Percentage depreciation = (0.0128 − 0.0133) / 0.0133 = −3.8%. (LOS 18.a)

Module Quiz 18.2

1. **C** This exchange rate regime is a currency board arrangement. The country has not formally dollarized because it continues to issue a domestic currency. A conventional fixed peg allows for a small degree of fluctuation around the target exchange rate. (LOS 18.b)

2. **B** An improvement in a trade deficit requires that domestic saving increase relative to domestic investment, which would decrease a capital account surplus. Decreasing expenditures relative to income means domestic savings increase. Decreasing domestic saving relative to domestic investment is consistent with a larger capital account surplus (an increase in net foreign borrowing) and a greater trade deficit. (LOS 18.b)

3. **B** Decreasing the volatility of domestic asset prices may be a goal of a government that imposes capital restrictions. Other typical goals include keeping domestic interest rates low and protecting certain domestic industries, such as the defense industry. (LOS 18.c)

READING 19

EXCHANGE RATE CALCULATIONS

MODULE 19.1: FOREIGN EXCHANGE RATES



Video covering this content is available online.

LOS 19.a: Calculate and interpret currency cross-rates.

A **cross rate** is the exchange rate between two currencies implied by their exchange rates with a common third currency. Cross-rates are necessary when there is no active FX market in a currency pair. The rate must be computed from the exchange rates between each of these two currencies and a third currency, usually the USD or EUR.

Let's assume that we have the following quotations for Mexican pesos and Australian dollars: MXN/USD = 10.70 and USD/AUD = 0.60. The cross rate between Australian dollars and pesos (MXN/AUD) is as follows:

$$\text{MXN/AUD} = \text{USD/AUD} \times \text{MXN/USD} = 0.60 \times 10.70 = 6.42$$

So, our MXN/AUD cross rate is 6.42 pesos per Australian dollar. The key to calculating cross-rates is to note that the basis of the quotations must be such that we get the desired result algebraically. If we had started with an AUD/USD quotation of 1.67, we would have taken the inverse to get the quotation into USD/AUD terms. Another approach is to divide through, as is illustrated in the following example.

EXAMPLE: Cross rate calculation

The spot exchange rate between the Swiss franc (CHF) and the USD is CHF/USD = 1.7799, and the spot exchange rate between the New Zealand dollar (NZD) and the U.S. dollar is NZD/USD = 2.2529. Calculate the CHF/NZD spot rate.

Answer:

The CHF/NZD cross rate is (CHF/USD) / (NZD/USD) = 1.7799 / 2.2529 = 0.7900.

LOS 19.b: Explain the arbitrage relationship between spot and forward exchange rates and interest rates, calculate a forward rate using points or in percentage

terms, and interpret a forward discount or premium.

When currencies are freely traded and forward currency contracts exist, the percentage difference between forward and spot exchange rates is approximately equal to the difference between the two countries' interest rates. This is because there is an arbitrage trade with a riskless profit to be made when this relation does not hold.

We call this a no-arbitrage condition because if it doesn't hold, there is an opportunity to make a profit without risk. The possible arbitrage is as follows: borrow Currency A at Interest Rate A, convert it to Currency B at the spot rate and invest it to earn Interest Rate B, and sell the proceeds from this investment forward at the forward rate to turn it back into Currency A. If the forward rate does not correctly reflect the difference between interest rates, such an arbitrage could generate a profit to the extent that the return from investing Currency B and converting it back to Currency A with a forward contract is greater than the cost of borrowing Currency A for the period.

The no-arbitrage condition requires that "you cannot earn more than your domestic riskless rate of interest by borrowing your domestic currency, converting it to a foreign currency to invest at the foreign riskless rate, and exchanging back to your domestic currency." So:

$$(1 + r_{\text{domestic}}) = \frac{1}{\text{spot}_{d/f}}(1 + r_{\text{foreign}})\,\text{forward}_{d/f}$$

The three terms on the right-hand side are equivalent to units of foreign currency for one unit of the domestic currency, growth if invested at the foreign interest rate, and then exchanged back to domestic currency at the forward rate.

Equivalent relationships are as follows:

$$\frac{\text{forward}_{d/f}}{\text{spot}_{d/f}} = \frac{(1 + r_{\text{domestic}})}{(1 + r_{\text{foreign}})}$$

$$\text{forward}_{d/f} = \frac{(1 + r_{\text{domestic}})}{(1 + r_{\text{foreign}})}\text{spot}_{d/f}$$

$$\text{spot}_{d/f} = \frac{(1 + r_{\text{foreign}})}{(1 + r_{\text{domestic}})}\text{forward}_{d/f}$$

EXAMPLE: Calculating the arbitrage-free forward exchange rate

Consider two currencies, the ABE and the DUB. The spot ABE/DUB exchange rate is 4.5671, the 1-year riskless ABE rate is 5%, and the 1-year riskless DUB rate is 3%. What is the 1-year no-arbitrage forward exchange rate?

Answer:

$$\text{forward}_{A/D} = \text{spot}_{A/D}\left(\frac{1 + r_A}{1 + r_D}\right)$$

$$\text{forward}_{A/D} = 4.5671\left(\frac{1.05}{1.03}\right) = 4.6558_{A/D}$$

Note that the forward rate is greater than the spot rate by 4.6558 / 4.5671 − 1 = 1.94%. This is approximately equal to the interest rate differential of 5% − 3% = 2%. The currency with the higher interest rate must depreciate over time by approximately the amount of the interest rate differential, to prevent arbitrage.

If we are calculating a 30-, 90-, or 180-day forward exchange rate, we need to use interest rates for 30-, 90-, and 180-day periods rather than annual rates. Note that these shorter-term rates are quoted as money market yields.

EXAMPLE: Calculating the arbitrage-free forward exchange rate with 90-day interest rates

The spot ABE/DUB exchange rate is 4.5671, the 90-day riskless ABE rate is 5%, and the 90-day riskless DUB rate is 3%. What is the 90-day forward exchange rate that will prevent arbitrage profits?

Answer:

$$\text{no-arbitrage forward} = 4.5671\left[\frac{1+0.05\left(\frac{90}{360}\right)}{1+0.03\left(\frac{90}{360}\right)}\right] = 4.5671\left(\frac{1.0125}{1.0075}\right)$$
$$= 4.5898 \text{ ABE/DUB}$$

In our previous example, we calculated the no-arbitrage one-year forward ABE/DUB exchange rate as 4.6558. Next, we illustrate the arbitrage profit that could be gained if the forward exchange rate differs from this no-arbitrage rate. Consider a forward rate of 4.6000 so that the depreciation in the ABE is less than that implied by the no-arbitrage relationship. This makes the ABE attractive to a DUB investor who can earn a riskless profit as follows:

- Borrow 1,000 DUB for one year at 3% to purchase ABE and get 4,567.1 ABE.
- Invest the 4,567.1 ABE at the ABE rate of 5% to have 1.05(4,567.1) = 4,795.45 ABE at the end of one year.
- Enter into a currency forward contract to exchange 4,795.45 ABE in one year at the forward rate of 4.6000 ABE/DUB to receive 4,795.45 / 4.6000 = 1,042.49 DUB.

The investor has ended the year with a 4.249% return on his 1,000 DUB investment, which is higher than the 3% 1-year DUB interest rate. After repaying the 1,000 DUB loan plus interest (1,030 DUB), the investor has a profit of 1,042.49 − 1,030 = 12.49 DUB with no risk and no initial out-of-pocket investment (i.e., a pure arbitrage profit).

Arbitrageurs will pursue this opportunity, buying ABE (driving down the spot ABE/DUB exchange rate) and selling ABE forward (driving up the forward ABE/DUB exchange rate), until the interest rate parity relation is restored and arbitrage profits are no longer available.

Note that the no-arbitrage forward exchange rate is approximately proportional to the annual interest rate differential and the time period of the forward contract.

Calculating a Forward Exchange Rate From a Forward Quote in Points or in Percentage Terms

A forward exchange rate quote typically differs from the spot quotation and is expressed in terms of the difference between the spot exchange rate and the forward exchange rate. One way to indicate this is with points. The unit of points is the last decimal place in the spot rate quote. For a spot currency quote to four decimal places, such as 2.3481, each point is 0.0001, or 1/10,000th. A quote of +18.3 points for a 90-day forward exchange rate means that the forward rate is 0.00183 greater than the spot exchange rate.

EXAMPLE: Forward exchange rates in points

The AUD/EUR spot exchange rate is 0.7313 with the 1-year forward rate quoted at +3.5 points. What is the 1-year forward AUD/EUR exchange rate?

Answer:

The forward exchange rate is 0.7313 + 0.00035 = 0.73165.

EXAMPLE: Forward exchange rates in percentage

The AUD/EUR spot rate is quoted at 0.7313, and the 120-day forward exchange rate is given as –0.062%. What is the 120-day forward AUD/EUR exchange rate?

Answer:

The forward exchange rate is 0.7313 (1 – 0.00062) = 0.7308.

While forward rates are typically quoted as forward points, one application where percentage forward quotes are useful is if we are interpreting a forward rate as the expected future spot rate, s_{t+1}. We can write our no-arbitrage relation as follows:

$$\frac{s_{t+1}}{s_t} - 1 = \%\Delta\, s_{t+1} = \frac{r_f - r_d}{1 + r_d}$$

Analysis of capital markets suggests that an increase in a country's interest rate will attract foreign investment, which will lead to appreciation of the domestic currency (spot rate decreases). This is opposite to our conclusion from our no-arbitrage forward rate result that higher domestic rates are associated with a depreciation of the domestic currency. Historically, interest rate differences are poor predictors of future spot rates, although they may be unbiased and tend to get the direction of change correct.

It's best to think of the no-arbitrage forward rate as simply the forward rate at a point in time that prevents currency arbitrage, rather than the expected future spot rate. Our no-arbitrage relation holds at a point in time and does not address the question of how changes in interest rates affect spot exchange rates over time.

Interpreting a Forward Discount or Premium

The **forward discount** or **forward premium** for a currency is calculated relative to the spot exchange rate. The forward discount or premium *for the base currency* is the percentage difference between the forward price and the spot price.

Consider the following spot and forward exchange rates:

USD/EUR spot = $1.312

USD/EUR 90-day forward = $1.320

The (90-day) forward premium or discount on the euro = forward / spot − 1 = 1.320 / 1.312 − 1 = 0.610%. Because this is positive, it is interpreted as a forward premium on the euro of 0.610%. Because we have the forward rate for three months, we could annualize the discount simply by multiplying by 4 (= 12 / 3).

Because the forward quote is greater than the spot quote, it will take more dollars to buy one euro 90 days from now, so the euro is expected to appreciate versus the dollar, and the dollar is expected to depreciate relative to the euro.

If the forward quote were less than the spot quote, the calculated amount would be negative, and we would interpret that as a forward discount for the euro relative to the U.S. dollar.



MODULE QUIZ 19.1

1. Today's spot rate for the Indonesian rupiah (IDR) is IDR/USD 2,400.00, and the New Zealand dollar trades at NZD/USD 1.6000. The NZD/IDR cross rate is:
   A. 0.00067.
   B. 1,492.53.
   C. 3,840.00.

2. The New Zealand dollar (NZD) is trading at USD/NZD 0.3500, and the Swedish krona (SEK) is trading at NZD/SEK 0.3100. The USD/SEK cross rate is:
   A. 0.1085.
   B. 8.8573.
   C. 9.2166.

3. The spot Swiss franc/British pound (CHF/GBP) exchange rate is 1.3050. In the 180-day forward market, the CHF/GBP exchange rate is −42.5 points. The 180-day forward CHF/GBP exchange rate is *closest* to:
   A. 1.2625.
   B. 1.3008.
   C. 1.3093.

4. The spot rate on the New Zealand dollar (NZD) is NZD/USD 1.4286, and the 180-day forward rate is NZD/USD 1.3889. This difference means:
   A. interest rates are lower in the United States than in New Zealand.
   B. interest rates are higher in the United States than in New Zealand.
   C. it takes more NZD to buy one USD in the forward market than in the spot market.

5. The current spot rate for the British pound (GBP) in terms of U.S. dollars is $1.533 and the 180-day forward rate is $1.508. Relative to the pound, the dollar is trading *closest* to a 180-day forward:
   A. discount of 1.63%.
   B. premium of 1.66%.

C. discount of 1.66%.

6. The annual interest rates in the United States (USD) and Sweden (SEK) are 4% and 7% per year, respectively. If the current spot rate is SEK/USD 9.5238, then the 1-year forward rate in SEK/USD is:
   A. 9.2568.
   B. 9.7985.
   C. 10.2884.

7. The annual risk-free interest rate is 10% in the United States (USD) and 4% in Switzerland (CHF), and the 1-year forward rate is USD/CHF 0.80. Today's USD/CHF spot rate is *closest* to:
   A. 0.7564.
   B. 0.8462.
   C. 0.8888.

KEY CONCEPTS

LOS 19.a

Given two exchange rate quotes for three different currencies, we can calculate a currency cross rate. If the MXN/USD quote is 12.1 and the USD/EUR quote is 1.42, we can calculate the cross rate of MXN/EUR as 12.1 × 1.42 = 17.18.

LOS 19.b

The condition that must be met so that there is no arbitrage opportunity available is as follows:

$$\frac{\text{forward}}{\text{spot}} = \frac{\left(1 + i_{\text{price currency}}\right)}{\left(1 + i_{\text{base currency}}\right)} \text{ so that forward} = \text{spot} \times \frac{\left(1 + i_{\text{price currency}}\right)}{\left(1 + i_{\text{base currency}}\right)}$$

If the spot exchange rate for the euro is 1.25 USD/EUR, the euro interest rate is 4% per year, and the dollar interest rate is 3% per year, the no-arbitrage one-year forward rate can be calculated as follows:

$$1.25 \times (1.03 \,/\, 1.04) = 1.238 \text{ USD/EUR}$$

Points in a foreign currency quotation are in units of the last digit of the quotation. For example, a forward quote of +25.3 when the USD/EUR spot exchange rate is 1.4158 means that the forward exchange rate is 1.4158 + 0.00253 = 1.41833 USD/EUR.

For a forward exchange rate quote given as a percentage, the percentage (change in the spot rate) is calculated as forward / spot − 1. A forward exchange rate quote of +1.787%, when the spot USD/EUR exchange rate is 1.4158, means that the forward exchange rate is 1.4158 (1 + 0.01787) = 1.4411 USD/EUR.

If a forward exchange rate does not correctly reflect the difference between the interest rates for two currencies, an arbitrage opportunity for a riskless profit exists. In this case, borrowing one currency, converting it to the other currency at the spot rate, investing the proceeds for the period, and converting the end-of-period amount back to the borrowed currency at the forward rate will produce more than enough to pay off the initial loan, with the remainder being a riskless profit on the arbitrage transaction.

To calculate a forward premium or discount for the base currency, use the following formula:

$$(\text{forward} / \text{spot}) - 1$$

Forward exchange rates are poor predictors of future exchange rates. Forward exchange rates should be interpreted as simply the no-arbitrage forward rates at a point in time, not as predictors of future spot rates.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 19.1

1. **A** Start with one NZD and exchange for 1 / 1.6 = 0.625 USD. Exchange the USD for 0.625 × 2,400 = 1,500 IDR. We get a cross rate of 1,500 IDR/NZD, or 1 / 1,500 = 0.00067 NZD/IDR. (LOS 19.a)

2. **A** USD/NZD 0.3500 × NZD/SEK 0.3100 = USD/SEK 0.1085. Notice that the NZD term cancels in the multiplication. (LOS 19.a)

3. **B** The 180-day forward exchange rate is 1.3050 − 0.00425 = CHF/GBP 1.30075. (LOS 19.b)

4. **B** Interest rates are higher in the United States than in New Zealand. It takes fewer NZD to buy one USD in the forward market than in the spot market. (LOS 19.b)

5. **B** To calculate a percentage forward premium or discount for the U.S. dollar, we use the dollar to be the base currency. The spot and forward quotes given are U.S. dollars per British pound (USD/GBP), so we must invert them to GBP/USD. The spot GBP/USD price is 1 / 1.533 = 0.6523 and the forward GBP/USD price is 1 / 1.508 = 0.6631. Because the forward price is greater than the spot price, we say the dollar is at a forward premium of 0.6631 / 0.6523 − 1 = 1.66%. Alternatively, we can calculate this premium with the given quotes as spot / forward − 1 to get 1.533 / 1.508 − 1 = 1.66%. (LOS 19.b)

6. **B** The forward rate in SEK/USD is $9.5238\left(\frac{1.07}{1.04}\right) = 9.7985$. Because the SEK interest rate is the higher of the two, the SEK must depreciate approximately 3%. (LOS 19.b)

7. **A** We can solve for the spot rate as follows. With the exchange rates quoted as USD/CHF, the spot is $0.80\left(\frac{1.04}{1.10}\right) = 0.7564$. Because the interest rate is higher in the United States, it should take fewer USD to buy CHF in the spot market. In other words, the USD must be at a forward discount relative to the CHF. (LOS 19.b)

TOPIC QUIZ: ECONOMICS

You have now finished the Economics topic section. Please log into your Schweser online dashboard and take the Topic Quiz on this section. The Topic Quiz provides immediate feedback on how effective your study has been for this material. Questions are more exam-like than typical Module Quiz or QBank questions; a score of less than 70% indicates that your study likely needs improvement. These tests are best taken timed; allow 1.5 minutes per question.

READING 20

ORGANIZATIONAL FORMS, CORPORATE ISSUER FEATURES, AND OWNERSHIP

MODULE 20.1: FEATURES OF CORPORATE ISSUERS



Video covering this content is available online.

LOS 20.a: Compare the organizational forms of businesses.

Organizational forms refer to how businesses are set up from a legal and organizational point of view. Key features of organizational forms include the following:

- Whether the business is a separate legal entity from the owner(s)
- Whether the owners of the business also operate the business, and if not, the nature of the relationship between its owners and operators
- Whether the owners' liability for the actions and debts of the business is limited or unlimited
- The tax treatment of profits or losses from the business
- Access to additional capital to fund expansion and to distribute risk

To understand these features, we can compare them among four commonly used types of business structures: sole proprietorships, general partnerships, limited partnerships, and corporations.

A **sole proprietorship** is a business owned and operated by an individual. Legally, the business is an extension of the owner, who is personally responsible for claims against the business (i.e., unlimited liability) and receives all profits/losses.

Profits are then taxed as personal income of the owner. Sole proprietorships tend to be small in scale because they can only expand within the limits of the individual owner's ability to secure financing.

To do business on a scale that exceeds that of a sole proprietorship, two or more individuals can form a **general partnership**. In this structure, the **partnership agreement** specifies each partner's responsibilities for business operations and their shares of the partnership profits or losses. The agreement may be written, verbal, or even incidental through the actions of the partners. As with a sole proprietorship, the

partners have unlimited liability for claims against the business, and profits from the business allocated to each partner are taxed as personal income.

A **limited partnership** involves two levels of partners. One or more general partners operate the business and have unlimited liability, as in a general partnership, but this structure also has **limited partners** who are liable only for the amount they invest in the partnership (i.e., **limited liability**) and have claims to its profits that are proportionate to their investments. Limited partners typically are not involved in appointing or removing general partners. How the profits are divided among the general and limited partners is specified in the partnership agreement. Because they are responsible for managing the business, the general partners typically receive a larger portion of profits than the limited partners. Profits allocated to the partners are taxed as personal income to each partner. Some jurisdictions allow a **limited liability partnership (LLP)** wherein a general partner is not required, and all the partners are limited partners. In the United States, LLPs are only allowed for providers of professional services such as law, accounting, and so on, and there are restrictions on the number of partners and the amount of equity investment.

The feature that distinguishes a **corporation**, or **limited company**, from the other business structures is that a corporation is a legal entity separate from its owners and managers. In this case, *all* of the corporation's shareholders have limited liability. An owner can lose his entire investment if the company goes bankrupt and the value of his shares goes to zero. Beyond their initial investments, shareholders are not responsible for claims against the corporation. A corporation may, but is not required to, distribute its profits to its owners. Most large firms are corporations because that structure gives them the greatest access to capital, both debt (**borrowed capital**) and equity (**ownership capital**).

Another distinguishing feature of corporations is the separation of its owners and managers. An investor who buys shares of a corporation does not directly influence the company's day-to-day operations. Instead, the owners appoint a **board of directors** that is responsible for hiring the senior managers to operate the company. The board and the managers it hires are responsible for acting in the interests of the shareholders.

A **public corporation** (or a **public limited company**) is one that has shares that are sold to the public and trade in an organized market (stock exchange). A **private limited company** is similar to a public company, but it has a limited number of shareholders and restrictions on transfer of shares.

Depending on the country, a corporation's profits may be subject to **double taxation** if the government taxes companies on their earnings *and* it taxes dividends (which are distributions of earnings to owners) as personal income.

LOS 20.b: Describe key features of corporate issuers.

A corporation's legal identity is separate from that of its owners and is formed by filing an **articles of incorporation** with a regulatory body. As a legal entity, a corporation has many of the rights and responsibilities of an individual, such as the right to hire employees, enter into contracts, borrow and lend money, and so on.

A corporation issues shares to the owners (shareholders), which allows it to raise large amounts of capital. Shareholders have **voting rights** that allow them to elect the board of directors. Shares are easily transferable if they are traded on an exchange. The board of directors may distribute a portion of the company's earnings to the shareholders as **dividends**.

The disadvantage of double taxation of corporate income is less for investors in companies that pay out a smaller fraction of their profits in dividends and reinvest the remaining, as illustrated in the following example.

EXAMPLE: Double taxation of dividends

ABC Corporation has pretax earnings of $10 million and a corporate tax rate of 25%. Shareholders are taxed at 20% on their dividend income. Calculate the effective tax rate if:

(a) 100% of the profits are paid out in dividends, and

(b) 40% of the profits are paid out in dividends.

Answer:

	(a) 100% payout	(b) 40% payout
Earnings before tax	$ 10,000,000	$ 10,000,000
(–) Corporate income tax @ 25%	2,500,000	2,500,000
(=) After-tax income	7,500,000	7,500,000
Dividends	7,500,000	3,000,000
Tax on dividends @ 20%	1,500,000	600,000
Total tax paid	4,000,000	3,100,000
Effective tax rate		
= Total tax / earnings before tax	40%	31%

LOS 20.c: Compare publicly and privately owned corporate issuers.

Most public limited companies are **listed companies**, which means their shares are listed on an exchange, allowing investors to trade shares. A stock exchange is a rules-based open market, providing price and volume transparency. Shareholders in a company can be individuals, other corporations, nonprofits, or government. Shares that are actively traded (i.e., not held by insiders, strategic investors, or sponsors) are called the company's **free float**. Free float is typically expressed as a percentage of total outstanding shares.

Public companies are subject to compliance and reporting requirements. For example, companies are required to file quarterly or annual financial reports with a regulatory body, and to disclose any material changes in the company's business or ownership.

Shares in private limited companies do not trade on an exchange. As a result, their value is not readily observable, and transfer between investors is difficult. Typically,

investors in a private limited company have to wait until the company goes public or is sold to exit their investment. Private companies have fewer regulatory requirements and typically disclose less information than public companies. Also, with fewer investors, private companies can take a longer-term view of the business.

Private companies can raise equity capital through **private placements** of securities. Private placements are typically restricted to **accredited investors** such as corporate and institutional investors or high net worth individuals.

Private companies can become public companies ("go public") in one of three ways: initial public offering, direct listing, or acquisition by a special purpose entity.

- A company can become public by issuing shares in an **initial public offering (IPO)**. To conduct an IPO, the company must meet exchange-specific requirements. Companies typically engage the services of an investment bank to underwrite the issue. Once the shares are listed on an exchange, owners can sell shares, and new owners can buy shares, without dealing directly with the company.
- In a **direct listing**, a stock exchange agrees to list a private company's existing shares. This differs from an IPO in that a direct listing does not raise any new capital for the company, but it has advantages in that it can be done more quickly than an IPO and without involving an underwriter.
- A **special purpose acquisition company (SPAC)** is a corporate structure set up to acquire a private company in the future. The SPAC raises capital through an IPO and puts the funds into a trust that it must use to make an acquisition within a specified time. The acquired company does not have to be identified at the time of the IPO. For this reason, SPACs are also known as "blank check" companies.

Sometimes there can be a benefit to taking an underperforming public company private to restructure it and unlock its potential value. In such a case, an acquirer purchases all the outstanding shares of a public company, and the company is delisted from the exchange. The benefits of going private include a lower regulatory burden and the associated cost savings.



MODULE QUIZ 20.1

1. Which organizational form has the highest degree of separation between the owners and operators of a business?
   A. Corporation.
   B. Limited partnership.
   C. General partnership.

2. A corporation is subject to a corporate income tax of 10%. Shareholders are taxed on dividends at a rate of 15%. Assume that the company pays out all its after-tax profits in dividends. The effective tax rate on the corporation's profit is:
   A. 12.5%.
   B. 23.5%.
   C. 25.0%.

3. A private company can become a public company through a:
   A. private placement.
   B. leveraged buyout.

C. special purpose acquisition company.

KEY CONCEPTS

LOS 20.a

Feature	Sole Proprietorship	General Partnership	Limited Partnership	Corporation
Separate legal entity?	No	No	No	Yes
Managed by	Owner	Partners	GPs	Board and managers
Owner liability	Unlimited	Unlimited	GPs: unlimited LPs: limited	Limited
Access to capital	Limited	Limited	Limited	Virtually unlimited
Taxation of profits	Pass through to owner as personal income	Pass through to owner as personal income	Pass through to owner as personal income	Potential double taxation

LOS 20.b

Corporations incorporate as a separate legal entity by registering with a regulatory body. Owners are shareholders with a limited liability, but profits may be subject to double taxation. Shareholders vote for a board of directors who then hire managers and declare dividend payments to the shareholders.

LOS 20.c

Public limited companies list their shares on a stock exchange for trading. Free float, or actively traded shares, is typically quoted as a fraction of total shares outstanding. Exchange listing requires regulatory compliance and extensive disclosure requirements.

Private companies raise capital in private placements. A private company can become public by issuing shares in an initial public offering, carrying out a direct listing on a stock exchange, or being acquired by a public company, which may be a special purpose acquisition company.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 20.1

1. **A** In a corporation, owners are most often not directly involved in operating the business. Both general partnerships and limited partnerships have general partners who operate the business. (LOS 20.a)

2. **B** The effective tax rate on profits distributed as dividends = 0.10 + 0.15 (1 − 0.10) = 23.5%. Alternatively:

- Starting with a hypothetical $100 of pretax profit, the corporate tax @10% is $10.
- Dividends = after-tax profit = $90. Tax on dividends @ 15% = 0.15 × 90 = $13.50.
- Total tax = 10 + 13.50 = $23.50, or 23.5% of the pretax profit of $100.

(LOS 20.b)

3. **C** An IPO, direct listing, and special purpose acquisition company are methods for a private company to go public. Private placements allow a private company to raise capital but not take it public. Buyouts can result in a public company going private. (LOS 20.c)

READING 21

INVESTORS AND OTHER STAKEHOLDERS

MODULE 21.1: STAKEHOLDERS AND ESG FACTORS



Video covering this content is available online.

LOS 21.a: Compare the financial claims and motivations of lenders and shareholders.

A company's lenders (debtholders) have a legal, contractual claim to the interest and principal payments the company has promised to make. Owners (equity holders) have a residual claim to the company's net assets (i.e., what remains after all other claims have been paid). That is, lenders have a higher priority of claims than equity owners. Because debt is less risky than equity, it is a less costly form of capital.

Both debtholders and equity holders can potentially lose their entire investment if a company fails, but their losses cannot exceed the amounts they have invested. A key difference between debt and equity investments is their upside potential. Regardless of a company's success, the best result debtholders can achieve is to receive the interest and principal payments promised by the company. Equity, on the other hand, has a theoretically unlimited upside if a company succeeds and grows over time.

The value of a company is the sum of the value of its debt and the value of its equity. As the company's value increases (assuming its value is greater than the value of its debt), the value of equity increases with it, while the value of debt is constant (i.e., no upside for debt investors from company growth). If the value of a company is less than the value of its debt, the value of its equity is zero, and the value of its debt is below the promised amount and moves with the value of the company (i.e., debt has downside exposure).

EXAMPLE: Impact of leverage on return on equity (ROE)

A company with revenues of $1,000 and operating expenses of $800 needs to invest $1,000 in assets. The assets can be 100% equity or a mix of 50/50 debt and equity. The interest rate on debt is 10% (assume no taxes). Calculate the ROE under both financing scenarios.

Answer:

	100% Equity	50% Debt, 50% Equity
Revenues	$1,000	$1,000
Cash operating expense	$800	$800
Interest expense	$0	$50
Net income	$200	$150
Equity	$1,000	$500
ROE	20%	30%

As can be seen in the previous example, increasing leverage (and therefore, risk) can increase the return on equity as long as the *expected* rate of return on assets exceeds the cost of debt. If things work out, and the actual return on assets is equal to or exceeds the cost of debt, the equity investors benefit from taking the risk. Debt investors, however, receive only their promised 10% interest (and the return of principal when the debt matures).

EXAMPLE: Impact of a decrease in revenues

Continuing the previous example, calculate the company's ROE if its revenues decrease by 15%, assuming no change in cash operating expenses.

Answer:

	100% Equity	50% Debt, 50% Equity
Revenues	$850	$850
Cash operating expense	$800	$800
Interest expense	$0	$50
Net income	$50	$0
Equity	$1,000	$500
ROE	5%	0%

Assuming no change in cash operating expenses, a 15% decrease in revenues would reduce the ROE to 5% (100% equity scenario) or to 0% (50/50 debt and equity scenario). If the decrease in revenues exceeded 15%, revenues would not have been enough to cover operating expenses, and debt investors would receive less than their promised 10% interest. Again, debt investors have no upside—only potential downside.

Because of the difference in their risk profiles, the interests of debtholders may conflict with the interests of equity holders. Debtholders are primarily concerned with a company's ability to repay its obligations and less concerned with its growth prospects. Equity holders may favor actions that increase a company's potential growth, but also increase its risk level, such as adding financial leverage by issuing new debt. Issuing

additional debt as opposed to equity also prevents **dilution** of shareholders' proportional ownership. A company's existing debtholders may oppose such actions because increasing the company's risk (and the probability of defaulting on its debts) does not increase their expected return. Therefore, debt investors usually limit borrowers' actions by including contractual provisions such as maximum leverage or a minimum interest coverage ratio (known as *covenants*, which we will examine in the Fixed Income topic area) in debt agreements.

LOS 21.b: Describe a company's stakeholder groups and compare their interests.

Under **shareholder theory**, the primary focus of corporate governance is the interests of the firm's shareholders, which is to maximize the market value of the firm's common equity. Under this theory, corporate governance is primarily concerned with the conflict of interest between the firm's managers and its owners (shareholders).

The focus of corporate governance under **stakeholder theory** is broader. It considers conflicts among several groups that have an interest in the activities and performance of the firm.

Stakeholders of a Corporation

Shareholders have a *residual interest* in the corporation, in that they have claim to the net assets of the corporation after all liabilities have been settled. Shareholders have voting rights to elect the board of directors and for other important corporate matters, which gives them control over the firm and its management. They have an interest in ongoing profitability and growth of the firm that will increase the value of their ownership shares.

Lenders can be public or private debtholders. Public debtholders or **bondholders** supply debt capital to the firm and are primarily owners of the firm's outstanding bonds, while **private debtholders** such as banks extend loans, credit facilities, and leases to the firm.

Private debtholders may have access to nonpublic information from company management, which decreases information asymmetry. For this reason, private debtholders are a critical source of financing for small-to-medium-sized businesses. In some cases, private debtholders hold equity in the firm, allowing them to take a more equity-like approach to evaluating the company and potentially making them more amenable to changes in terms of the loans (i.e., covenants). By contrast, bondholders rely only on public information and have little to no influence over an issuer's operations. The interests of both types of lenders are protected to varying degrees by covenants in their debt agreements with the firm.

The **board of directors** is responsible for protecting the interests of shareholders; hiring, firing, and setting the compensation of the firm's senior managers; establishing the strategic direction of the firm; and monitoring the company's financial performance and other aspects of its ongoing activities.

Board members include **inside directors** (e.g., senior executives, founders) and **independent directors** who have no material relationship with the company. While inside directors may have conflicts of interest with shareholders, independent directors may better protect shareholders' interests.

In a **one-tier board structure**, both inside and independent directors serve on a single board. Major stock exchanges specify requirements for director independence, such as that most of the board should comprise independent directors. Other requirements may include diversity of backgrounds and competencies. In continental Europe, boards have a **two-tier structure** in which the independent directors serve on a **supervisory board** that oversees a **management board** comprising inside directors.

Typically, board members are elected for a specified term (e.g., annual). In a **staggered board**, only a fraction of the board is elected each year. This decreases the power of shareholders to enact a major overhaul of the board. Firms with staggered boards justify them as providing continuity and allowing for a longer-term view of company strategy.

Senior managers typically receive compensation (remuneration) that is made up of a salary, a bonus based on some measure of company performance, and perks (e.g., expense accounts, use of company planes, special retirement benefits). Their interests can be expected to include continued employment and maximizing the total value of their compensation. Executive bonuses are typically tied to some measure of firm performance, giving senior managers a strong interest in the financial success of the firm.

Other employees are the **human capital** of the company, who provide their labor and skills. Employees also have an interest in the sustainability and success of the firm. They have interests in their rate of pay, opportunities for career advancement, training, and working conditions. Employees may hold equity in the company by participating in employee stock participation plans, which are designed to align their interests with those of the shareholders of the company. In some industries, employees join unions to negotiate the terms of their employment.

Suppliers of resources to the firm have an interest in preserving an ongoing relationship with the firm, in the profitability of their trade with the firm, and in the growth and ongoing stability of the firm. Because suppliers are typically short-term creditors of the firm, they also have an interest in the firm's solvency and ongoing financial strength.

Customers rely on the company to provide a good or service of high quality at a reasonable price. Customers may also have an interest in preserving an ongoing relationship to receive support and after-sale service. Maintaining the good will of customers is critical for the long-term prospects of any company. Customers increasingly care about environmental and social responsibilities of corporations.

Governments rely on corporations for tax revenue, economic growth, social welfare, and employment creation. Regulators have an interest in ensuring compliance with various laws.

LOS 21.c: Describe environmental, social, and governance factors of corporate issuers considered by investors.

Some debt and equity investors who take a stakeholder perspective are interested in evaluating companies' environmental, social, and governance (ESG) factors. This is primarily due to three reasons:

1. Government stakeholders increasingly prioritize climate change and social policies through regulatory changes.
2. ESG factors can have a material impact on companies' results through potential loss of customer good will and financial losses due to fines and judgments. Poor corporate governance may lead to senior managers exploiting shareholders to advance personal interests.
3. Many younger investors increasingly manage their wealth with ESG considerations in mind.

Negative externalities arise when a company or its investors do not bear the full cost of its actions (e.g., environmental damage). Increased government regulations and stakeholder awareness necessitate companies to recognize these costs, either explicitly in their financial statements or implicitly.

Environmental Factors

Environmental factors include company contributions to problems such as climate change, air and water pollution, deforestation, energy efficiency, waste management, or water scarcity. Material environmental factors can have a substantial impact on companies' operations or business models. Industries that are natural resource intensive have a direct impact on the environment, but other industries may have indirect impacts.

With respect to climate change, companies may face **physical risk** of adverse effects on assets or operations if severe weather increases in frequency, as well as **transition risk** as government regulations or consumer choices require switching from high-carbon to low-carbon activities. **Stranded assets** are those that become unviable due to such changes. Poor safety policies or inadequate governance systems increase the risks for adverse events such as oil spills and contamination of groundwater. The costs of penalties, cleanup, litigation, and loss of reputation can be significant risks for a company's investors.

Social Factors

Social factors include the protection of customer privacy and information security, customer satisfaction, employee engagement, diversity and inclusion, labor relations, and community relations. Social factors contribute to a company's image in terms of how it treats its employees, customers, and the communities in which it operates. Taking measures to decrease social risk can reduce a company's costs through higher employee productivity, lower turnover, increased customer loyalty, and less risk of litigation.

Governance

Corporate governance encompasses factors such as the composition of the board and the internal audit committee, executive compensation, bribery and corruption, political contributions, and lobbying. Corporate governance systems should have adequate checks to ensure that managers act ethically, lawfully, and in the interests of shareholders.

Evaluation of ESG Risks

Analysts should identify and measure the ESG-related risks to which a company may be exposed and how they can affect the company's future cash flows. Equity investors bear the brunt of the risk from adverse outcomes. Debt investors have less exposure to adverse events unless they result in losses large enough to bring about default. Not all debtholders are equally exposed to ESG risks. Because some such risks may be delayed (e.g., a coal-fired electricity plant that is currently in compliance may become obsolete in the future), longer-maturity debt investors may have more exposure than short-term debt investors.



MODULE QUIZ 21.1

1. For a company that is financially sound, increasing the company's rate of growth is *most likely* to benefit:
   A. equity holders.
   B. debtholders.
   C. neither debtholders nor equity holders.

2. Which of the following board structures is *most likely* to be preferred by a minority shareholder?
   A. Majority independent and staggered elections.
   B. Majority independent and full board election.
   C. Majority inside and staggered elections.

3. A company decides to shut down a production plant rather than retrofit it to comply with new environmental regulations. This is *best* described as an example of:
   A. governance factors.
   B. social factors.
   C. stranded assets.

KEY CONCEPTS

LOS 21.a

A company's debtholders have a higher priority of claims than its equity holders. Debtholders have a legal claim to the promised interest and principal payments. Equity holders have a residual claim to the company's net assets after debtholders have been paid.

Debt has limited upside potential because the best result for debtholders is to receive the promised principal and interest payments. Equity has theoretically unlimited upside potential. This difference may create conflicts of interest between debtholders and equity holders.

LOS 21.b

The primary stakeholders of a corporation include shareholders, debtholders, the board of directors, senior management, employees, creditors, suppliers, and government. Stakeholder theory postulates that a company needs to balance the interests of all stakeholders.

LOS 21.c

Environmental factors include company contributions to climate change, air and water pollution, deforestation, energy efficiency, waste management, and water scarcity.

Social factors include the protection of customer privacy and information security, customer satisfaction, employee engagement, diversity and inclusion, labor relations, and community relations.

Corporate governance factors include the composition of the board, executive compensation, the internal audit function, bribery and corruption, political contributions, and lobbying.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 21.1

1. **A** If a company is financially sound, it is repaying interest and principal in full and on time. Debtholders have no additional claims to increased company profits. Equity holders benefit from the upside of a company's growth. (LOS 21.a)

2. **B** Minority shareholders prefer independent directors, as they would act in the best interests of the shareholders as opposed to acting in the interests of the management. Full board election would allow the shareholders to vote out the board if it was ineffective. Staggered elections delay changes to the board membership. (LOS 21.b)

3. **C** Stranded assets arise from obsolescence of existing assets that do not conform to new environmental standards. (LOS 21.c)

READING 22

CORPORATE GOVERNANCE: CONFLICTS, MECHANISMS, RISKS, AND BENEFITS

MODULE 22.1: CORPORATE GOVERNANCE



Video covering this content is available online.

LOS 22.a: Describe the principal-agent relationship and conflicts that may arise between stakeholder groups.

When one party hires another to carry out a task, the two parties are said to have a **principal-agent relationship**. In such a relationship a **principal-agent conflict** has the potential to arise, because an agent is hired to act in the interests of the principal, but an agent's interests might not coincide exactly with those of the principal.

For example, consider an insurance agent who is paid a commission on policies written. It would be in the agent's interest to write insurance policies on people or property that are not good risks to maximize commission income. The principal (the owner of the insurance company) does not want to issue policies that are bad risks, as that is a money-losing proposition. Insurance companies mitigate this conflict by imposing underwriting standards for the policies they will issue and by continuing to work only with agents who consistently act in the company's best interest.

Agency costs are the costs of a principal-agent conflict. Agency costs can be direct, such as hiring employees to monitor the agent, or indirect, such as the opportunity cost of lost business.

Conflicts of Interest Between Shareholders and Managers or Directors

In the context of a corporation, shareholders are the principals (owners), and firm management and board members (directors) are their agents. Managers and directors may choose a lower level of business risk than shareholders would. Managers have their employment income tied to the firm and stand to lose it, should the firm fail. Shareholders can diversify their portfolios cheaply, and as such might prefer that individual companies take more risk.

Conflicts may arise if inside directors favor management interests at the expense of shareholder interests, or if directors favor one group of shareholders at the expense of

another. Some conflicts result from **information asymmetry** between shareholders and managers, which refers to the fact that managers have more and better information about the functioning of the firm and its strategic direction than shareholders do. This decreases the ability of shareholders or nonexecutive directors to monitor and evaluate whether managers are acting in the best interests of shareholders. Information asymmetry is more acute for larger companies that operate in many businesses and geographical markets, companies that sell complex products, and companies with lower levels of institutional ownership and free float.

Common principal-agent conflicts arise in the following ways:

- Managers may put in insufficient effort. This leads to poor evaluation of investment opportunities and risks taken, resulting in overall higher total costs.
- Managers getting option grants may want to ratchet up risk because options do not have any downside risk. On the other hand, managers who are compensated primarily with cash may be inadequately motivated to take risk.
- Manager compensation tied to company size creates incentives for *empire building* through poor or unnecessary acquisitions.
- Managers may seek to entrench themselves by taking inadequate risk, mimicking competitors' actions rather than generating original ideas, or engaging in otherwise unprofitable projects that require their specific individual knowledge. Directors, in turn, might seek to entrench themselves by going along with management's decisions instead of questioning them.
- Managers may undertake **self-dealing** by exploiting firm resources for personal benefit.

Conflicts Between Groups of Shareholders

A single shareholder or group of shareholders in a company may hold most of the votes (**controlling shareholders**) and act against the interests of the **minority shareholders**. A controlling shareholder may have concentrated ownership, in that a large proportion of his or her wealth might consist of shares in this company. In this situation a controlling shareholder might want the company to diversify into different businesses to mitigate risk. Minority shareholders, by contrast, might already hold diversified portfolios and do not want the company to squander resources by investing in a less desirable business just to diversify.

Some firms have a **dual-class structure** with different classes of common stock outstanding, some with more voting power than others. This can give a group of shareholders (e.g., the company's founders) effective control of the company even if they have claims to less than 50% of its earnings and assets. CFA Institute advocates against dual-class voting structures because they allow one group of shareholders to further their interests at the expense of other groups.

Conflicts of Interest Between Creditors and Shareholders

Shareholders may prefer more business risk than creditors do because creditors have a limited upside from good results compared to shareholders. Management actions that favor the interests of equity owners over the interests of creditors include issuing new debt that increases the default risk faced by existing debtholders, or increasing dividends at the cost of decreasing company assets as collateral and increasing the risk of default. This potential for conflict is a greater risk for long-term debtholders.

LOS 22.b: Describe corporate governance and mechanisms to manage stakeholder relationships and mitigate associated risks.

Corporate governance is the system of internal controls and procedures by which individual companies are managed. It includes a framework that defines the rights, roles, and responsibilities of various groups within an organization. The objective is to manage and minimize conflicts of interest between stakeholders of the company.

Stakeholder management refers to the management of company relations with stakeholders and is based on having a good understanding of stakeholder interests and maintaining effective communication with stakeholders. With respect to the company's relationship with shareholders, there are standard practices. These practices are required by corporate laws and are similar in many jurisdictions, although there are some differences across countries.

Stakeholders gather information about the company from public reports filed by the company. Public companies release information in annual reports, proxy statements, and public notices. This reporting includes information about the financial performance and standing of the company, related-party transactions, executive remuneration, and governance structure. While reporting requirements for private companies are more limited, these companies typically provide information to their investors directly. Transparency in reporting reduces information asymmetry and allows stakeholders to evaluate whether the company's actions align with their interests.

Shareholder Mechanisms

Corporations typically hold an **annual general meeting** after the end of the firm's fiscal year. At the general meeting, company management provides shareholders with the audited financial statements for the year, addresses the company's performance and significant actions over the period, and answers shareholder questions.

Corporate laws dictate when the annual general meeting must occur and how the meeting must be communicated to shareholders. Typically, anyone owning shares is permitted to attend the annual general meeting, to speak or ask questions, and to vote their shares. A shareholder who does not attend the annual general meeting can vote her shares by **proxy**, meaning she assigns her right to vote to another person who will attend the meeting—often a director, a member of management, or the shareholder's

investment advisor. A proxy may specify the shareholder's vote on specific issues or leave the vote to the discretion of the person to whom the proxy is assigned.

Ordinary resolutions, such as the approval of an auditor and the election of directors, require a simple majority of the votes cast. Other resolutions are addressed at **extraordinary general meetings**, which can be called any time a matter requires a shareholder vote. Examples include amendments to the company's bylaws, a merger or takeover, a shareholder-proposed special board election, or the liquidation of the firm.

Activist shareholders pressure companies in which they hold a significant number of shares for changes they believe will increase shareholder value. They may initiate shareholder lawsuits or seek representation on the board of directors. Other activist tactics include proposing shareholder resolutions for a vote and raising their issues to all shareholders or the public to gain wider support. Hedge funds have engaged increasingly in shareholder activism at firms in which they hold significant stakes.

A shareholder activist group may initiate a **proxy contest**, in which they seek the proxies of shareholders to vote in favor of their alternative proposals, or may make a **tender offer** for enough shares of a company to gain control. Both senior managers and boards of directors can be replaced by shareholders when they believe company performance is poor and would be improved by change. The threat of a **hostile takeover**, one not supported by the company's management, can act as an incentive for company managements and boards to pursue policies better aligned with the interests of shareholders. On the other hand, it might also cause the current management or board to adopt takeover defenses such as staggered board elections or **poison pill** provisions (low-price additional share offerings to current shareholders).

Creditor Mechanisms

When a company issues a bond, it specifies the rights of bondholders and the company's obligations in a legal document called a **bond indenture**. An indenture typically includes **covenants** that may require the company to take certain actions, or restrict it from taking certain actions. A bond can be backed by **collateral**, which is a specific asset pledged against which the bondholders will have a claim if the company defaults on the bond. A financial institution may act as a trustee to monitor the company's compliance with its bond covenants.

Creditor committees may form among bondholders to protect their interests when an issuer experiences financial distress. Some countries require such committees when a company files for bankruptcy. A group of bond investors may form an **ad hoc committee** when a company is struggling to meet its obligations. While the committee does not represent *all* the bondholders, their interests are generally aligned.

Board of Directors and Management Mechanisms

The **board of directors** is elected by shareholders to act in their interests. A board of directors typically has committees made up of board members with particular expertise. These committees report to the board, which retains the overall

responsibility for the various board functions. The following are examples of typical board committees.

An **audit committee** is responsible for the following:

- Oversight of the financial reporting function and implementation of accounting policies
- Effectiveness of the company's internal controls and the internal audit function
- Recommending an independent external auditor and its compensation
- Proposing remedies based on their review of internal and external audits

A **nominating/governance committee** is responsible for the following:

- Oversight of the company's corporate governance code including board elections
- Setting policies for nomination of candidates for board membership
- Implementing the company's code of ethics and policies regarding conflicts of interest
- Monitoring changes in relevant laws and regulations
- Ensuring that the company is in compliance with all applicable laws and regulations, as well as with the company's governance policies

A **compensation committee** or **remuneration committee** recommends to the board the amounts and types of compensation to be paid to directors and senior managers. This committee may also be responsible for oversight of employee benefit plans and evaluation of senior managers. Because managers should not be in a position to evaluate or compensate themselves, a compensation committee should (and in many countries is required to) be composed of independent directors only.

Other committees are industry specific. A risk committee (financial services industry) informs the board about appropriate risk policy and risk tolerance of the organization, and it oversees the enterprise-wide risk management processes of the organization. An investment committee (insurance industry) reviews and reports to the board prudent investment and capital management policies.

The number and size of board committees will depend on the size, complexity, and nature of the business. Regulations often require that firms have audit committees. The composition of a board committee is often based on its function, with audit committees, compensation committees, and governance committees often made up of only nonexecutive or independent directors.

Employee, Customer, and Supplier Mechanisms

Labor laws, employment contracts, and the right to form unions are the primary mechanisms for employees to manage relationships with employers. Some countries have laws that require boards of large companies to include employee representatives. **Employee stock ownership plans (ESOPs)** may help align company and employee interests. For customers and suppliers, contracts tend to be the mechanism through which they manage their relationships with companies. In recent years, customers and

other stakeholders have increasingly used social media as a mechanism to influence company behavior.

Government Mechanisms

Governments enact and enforce regulations that govern companies' actions. They often do so by establishing agencies to regulate industries or sectors such as financial markets, or by monitoring specific issues such as workplace safety and environmental protection. In some countries, regulators develop corporate governance codes that companies must either adopt or explain why they have not done so. In some countries, corporate governance regulations are specified by stock exchanges as part of their listing requirements.

LOS 22.c: Describe potential risks of poor corporate governance and stakeholder management and benefits of effective corporate governance and stakeholder management.

Risks of Poor Governance and Stakeholder Management

When corporate governance is weak, the control functions of audits and board oversight may be weak as well. The risk is that some stakeholders can gain an advantage, to the disadvantage of other stakeholders. Accounting fraud, or simply poor recordkeeping, will have negative implications for company performance and value.

When governance is weak and managers are not monitored, they may serve their own interests by choosing less-than-optimal risk, reducing company value. Without proper monitoring and oversight, management may be given incentive compensation that allows them to pursue their own benefit rather than the company's interests. They may engage in related-party transactions that benefit their friends or family, to the detriment of shareholders.

Poor compliance procedures with respect to regulation and reporting can easily lead to legal and reputational risks. Violating stakeholder rights can lead to stakeholder lawsuits. Failure to comply with government regulations can damage company reputation. Failure to manage creditors' rights well can lead to debt default and bankruptcy.

Benefits of Effective Governance and Stakeholder Management

Effective corporate governance can improve operational efficiency by ensuring that management and board member incentives align their interests well with those of the shareholders. Effective governance implies effective control and monitoring. Just as weak control can lead to abuses, a strong system of controls and compliance with laws and regulations can avoid many legal and regulatory risks.

Formal policies regarding conflicts of interest and related-party transactions can also lead to better operating results. Proper governance with respect to the interests of creditors can reduce the risk of debt default or bankruptcy, thereby reducing the cost of debt financing. Alignment of management interests with those of shareholders leads to better financial performance and greater company value.



MODULE QUIZ 22.1

1. Between which two of a company's stakeholder groups does information asymmetry *most likely* make monitoring more difficult?
   A. Suppliers and customers.
   B. Employees and government.
   C. Managers and shareholders.
2. The theory that deals with conflicts of interest between a company's owners and its creditors is *most appropriately* called:
   A. governance theory.
   B. stakeholder theory.
   C. shareholder theory.
3. Benefits of effective corporate governance and stakeholder management *most likely* include:
   A. reduced risk of default.
   B. more efficient related-party transactions.
   C. greater control exercised by the most interested stakeholders.

KEY CONCEPTS

LOS 22.a

The principal-agent relationship refers to owners employing agents to act in their interests. Conflicts can arise because the agent's incentives may not align with those of the owner or, more generally, because the interests of one group within a corporation are not the same as those of other groups.

LOS 22.b

Corporate governance refers to the internal controls and procedures of a company that delineate the rights and responsibilities of various groups, and how conflicts of interest among the various groups are to be resolved.

Shareholders, creditors, boards of directors, employees, customers, suppliers, and governments have different mechanisms with which to manage their stakeholder relationships with companies.

Proxy voting is the primary shareholder mechanism. Shareholders can remove senior managers and boards of directors if they believe company performance would improve with a change. Activist shareholders may engage in proxy fights or hostile takeovers.

Creditor mechanisms include bond indentures and creditor committees. Employee mechanisms include labor laws and unions. Contracts are the primary mechanism for customers and suppliers. Governments may enact regulations or appoint regulatory agencies.

Duties of a board of directors include the following:

- Selecting senior management, setting their compensation, and evaluating their performance
- Setting the strategic direction for the company
- Approving capital structure changes, significant acquisitions, and large investment expenditures
- Reviewing company performance and implementing any necessary corrective steps
- Planning for continuity of management and the succession of the CEO
- Establishing, monitoring, and overseeing the firm's internal controls and risk management
- Ensuring the quality of the firm's financial reporting and internal audit

LOS 22.c

The risks of poor governance include weak control systems, poor decision-making, legal risk, reputational risk, and default risk. Good corporate governance can improve operational efficiency and performance, reduce default risk, reduce the cost of debt, improve financial performance, and increase firm value.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 22.1

1. **C** Information asymmetry can exist between a company's shareholders and its managers because the company's managers may be much more knowledgeable about the company's functioning and strategic direction. This makes it more difficult for shareholders to monitor the firm's managers and determine whether they are acting in shareholders' interests. (LOS 22.a)

2. **B** Stakeholder theory focuses on the conflicts of interest among owners and several groups that have an interest in a company's activities, including creditors. (LOS 22.b)

3. **A** Reduced risk of default is among the benefits of effective corporate governance. Risks from poor corporate governance include related-party transactions by managers and opportunities for some stakeholder groups to gain an advantage at the expense of others. (LOS 22.c)

READING 23

WORKING CAPITAL AND LIQUIDITY

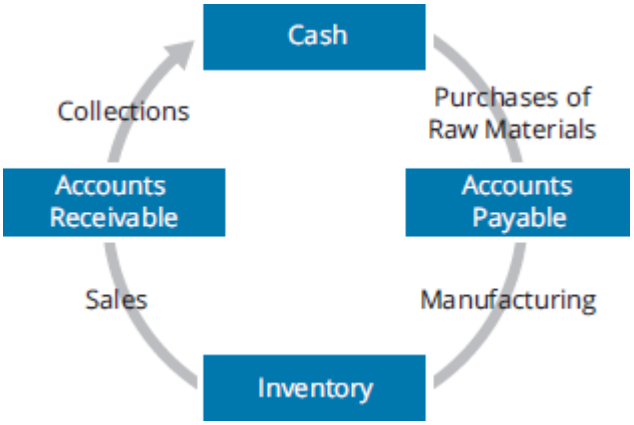
MODULE 23.1: LIQUIDITY MEASURES AND MANAGEMENT

Video covering this content is available online.

LOS 23.a: Explain the cash conversion cycle and compare issuers' cash conversion cycles.

The **cash conversion cycle (CCC)** measures the efficiency of a company's cash flow management. It represents the time it takes for a company to convert its investments in inventory and other resources into cash inflows from sales. In other words, it measures how quickly a company can convert its investments into cash and then use that cash to invest in new opportunities, as shown in Figure 23.1.

Figure 23.1: Cash Conversion Cycle



The CCC is calculated by adding the number of days it takes for a company to sell its inventory (**days of inventory on hand**), the number of days it takes for the company to collect payment from its customers (**days sales outstanding**), and the number of days it takes for the company to pay its suppliers (**days payable outstanding**).

$$\text{cash conversion cycle} = \text{days of inventory on hand} + \text{days sales outstanding} - \text{days payables outstanding}$$

PROFESSOR'S NOTE

The calculation of DOH, DSO, and DPO is covered in our Financial Statement Analysis reading on Financial Analysis Techniques.

A lower CCC is generally better as it indicates that a company can generate cash quickly and efficiently, meaning that less of the company's capital is devoted to working capital. A high CCC may indicate that a company is taking longer to convert its investments into cash, which could lead to cash flow problems and potentially limit the company's ability to invest in growth opportunities.

A company can decrease its CCC by decreasing its inventories and receivables or by increasing (extending) its payables. However, each of these has potential disadvantages:

- Reduced inventories of raw materials may create production bottlenecks due to supply chain disruptions. Reduced inventories of finished foods may mean an inability to meet spikes in customer demand.
- Reducing or tightening credit to customers may result in lost sales.
- We can think of accounts payable as an implicit source of credit from suppliers (as opposed to explicit sources such as bank loans). Suppliers offer payment terms in the form *a/b net c*, which means a percentage discount of *a* if the invoice is paid within *b* days, otherwise full payment is due within *c* days. Forgoing the discount for prompt payment amounts to borrowing money from the supplier for (c – b) days, at an effective annual rate we can calculate as follows:

  $$\text{EAR of supplier financing} = \left(1 + \frac{a}{1-a}\right)^{\left(\frac{365}{c-b}\right)} - 1$$

  where:
  a = percent discount
  b = days until discount expires
  c = days until full payment is due

  Often, the company could find a lower EAR by borrowing from their lenders as opposed to forgoing the discount.

EXAMPLE: EAR of supplier financing

A supplier offers 2/10 net 30 terms. The bank interest rate is 8%. Which source of financing should the company prefer?

Answer:

Financing is for 30 − 10 = 20 days.

EAR of supplier financing = $(1 + 0.02 / 0.98)^{(365/20)} - 1 = 0.446$, or 44.6%.

The cost of implicit supplier financing is much higher than the cost of explicit bank financing. The company should borrow from its bank at 8% to pay the invoice within 10 days.

Cash conversion cycles vary by industry. For example, pharmaceutical companies have long CCCs because they maintain inventories of high-margin drugs to meet unexpected surges in demand. By contrast, airlines have low CCCs because most of their sales are prepaid and they do not keep significant inventories. Analysts should use the CCC to

compare companies within the same industry, or to track a company's performance over time.

In addition to the CCC, overall levels of working capital can indicate how efficiently a company manages its liquidity. To compare companies of different sizes, analysts calculate working capital as a proportion of sales. Because the ratio varies by industry, care should be taken to compare firms in similar lines of business. **Total working capital** is current assets less current liabilities. Analysts often prefer to measure **net working capital**, using only operating current assets and operating current liabilities, because it is closely linked to the CCC:

$$\text{total working capital} = \text{current assets} - \text{current liabilities}$$

$$\text{net working capital} = \text{current assets (except cash and marketable securities)} - \text{current liabilities (excluding short-term and current debt)}$$

LOS 23.b: Explain liquidity and compare issuers' liquidity levels.

For an asset, **liquidity** refers to its nearness to cash. For a liability, liquidity refers to its nearness to settlement. Assets that can be quickly converted to cash (e.g., marketable securities) are considered highly liquid.

Inventory is less liquid than accounts receivable. Inventory may need processing before a sale, and once sold, inventory might be converted to accounts receivable, which needs to be collected to convert into cash.

For a corporate issuer, liquidity refers to availability of cash and other liquid assets to meet its short-term obligations. **Primary liquidity sources** include cash and marketable securities on hand, bank borrowings, and cash generated from the business. Long-term solvency of a company depends on its ability to generate sufficient cash from its business to service its liabilities. Analysts evaluate a company's liquidity management primarily by reviewing its statement of cash flows.

Companies generally rely on primary sources of liquidity. However, if needed, the company can rely on **secondary liquidity sources**, such as the following:

- Cash saved by suspending dividends to shareholders
- Delaying or reducing capital investments
- Selling assets
- Issuing additional equity
- Restructuring debt to extend its maturity
- Bankruptcy protection filing, which suspends the need to service the liabilities

Using secondary liquidity sources sends negative signals to the market, and they are generally more costly than primary liquidity sources.

A company's cash conversion cycle may vary by time of the year (i.e., seasonality) and as business conditions change (e.g., a sudden dip in demand for a company's product will increase inventory and increase DOH). Cash and marketable securities serve as a buffer to meet the company's obligations when there is a deviation from the normal

CCC. This excess cash has a cost because it represents capital not invested in the business, but inadequate liquidity can result in having to rely on higher-cost secondary sources.

EXAMPLE: Cost of liquidity

Drake Corporation has a sudden need for liquidity to meet its payroll at the end of the week. The company's sources of liquidity and their costs are as shown here.

Source	$ Fair Market Value ($000s)	Liquidation Cost (%)
Cash and marketable securities	100	0
Inventory and receivables	200	15
Empty warehouse	300	30
Total	600	

Calculate the total proceeds from all sources, and the cost of liquidity in dollars and in percentage.

Answer:

The cost of liquidity is the discount to fair market value for which an asset is sold due to the need for immediate liquidity, rather than waiting for the normal time it takes to sell the asset.

Source	Fair Market Value ($000s)	Liquidation Cost		Net Proceeds
Cash and marketable securities	$100	0%	$0	$100
Inventory and receivables	$200	15%	$30	$170
Empty warehouse	$300	30%	$90	$210
Total	$600		$120	$480

The cost of liquidity = $120,000, or 120 / 600 = 20%.

Factors Affecting Liquidity

An increase in the CCC reduces an issuer's liquidity. The CCC can increase due to drags or pulls on liquidity. A **drag on liquidity** occurs when cash inflows lag. This can occur when excess inventory builds up or inventory becomes obsolete (DOH increases), or when collections are slow or receivables become uncollectible (DSO increases). A **pull on liquidity** occurs when cash outflows accelerate. This can occur when suppliers reduce credit lines or demand faster payments (DPO decreases).

Liquidity Ratios

The **current ratio** is the ratio of current assets to current liabilities. A current ratio greater than 1 indicates that the company has sufficient current assets to meet its current liabilities.

$$\text{current ratio} = \frac{\text{current assets}}{\text{current liabilities}}$$

Not all current assets are equally liquid. Because inventory is the least liquid current asset, the **quick ratio** excludes inventory from current assets and is a more stringent metric of liquidity.

$$\text{quick ratio} = \frac{(\text{cash and marketable securities} + \text{accounts receivable})}{\text{current liabilities}}$$

Finally, the **cash ratio** excludes both inventory and accounts receivables and is the most stringent measure of liquidity.

$$\text{cash ratio} = \frac{\text{cash and marketable securities}}{\text{current liabilities}}$$

EXAMPLE: Liquidity ratios

The following information is available for Drake Corporation for the past two years:

Current Assets	20X1 ($000s)	20X2 ($000s)
Cash and marketable securities	$ 98	$ 54
Accounts receivable	$ 120	$ 110
Inventory	$ 190	$ 310
Prepaid accounts	$ 24	$ 18
Current Liabilities		
Accounts payable	$ 85	$ 100
Accrued expenses	$ 34	$ 39
Short-term notes payable	$ 98	$ 145

Calculate Drake's current ratio, quick ratio, and cash ratio for each of the two years and comment on the trend in the ratios.

Answer:

Current ratio = current assets / current liabilities

Current ratio (20X1) = (98 + 120 + 190 + 24) / (85 + 34 + 98) = 1.99

Current ratio (20X2) = (54 + 110 + 310 + 18) / (100 + 39 + 145) = 1.73

Quick ratio = cash and marketable securities + accounts receivable / current liabilities

Quick ratio (20X1) = (98 + 120) / (85 + 34 + 98) = 1.00

Quick ratio (20X1) = (54 + 110) / (100 + 39 + 145) = 0.58

Cash ratio = cash and marketable securities / current liabilities

Cash ratio (20X1) = 98 / (85 + 34 + 98) = 0.45

Cash ratio (20X2) = 54 / (100 + 39 + 145) = 0.19

All three ratios are declining year-over-year from 20X1 to 20X2, indicating that Drake's liquidity is worsening.

LOS 23.c: Describe issuers' objectives and compare methods for managing working capital and liquidity.

Working capital management seeks to maximize firm profits while ensuring that sufficient liquidity is available to maintain the firm's operations and meet its obligations. Firms may choose to hold more short-term assets (which offer lower rates of return) to ensure that sufficient cash is available to service its obligations. The firm can also choose to finance its working capital using short-term loans, which are cheaper than sources such as long-term debt and equity. However, the firm must consider the risk of being able to roll over short-term debt at a reasonable cost.

In analyzing different approaches to working capital management, it is important to control for size differences between companies. We do this by calculating companies' relative amounts of long-term versus short-term assets as a proportion of sales.

A conservative approach to working capital management is for the company to hold higher amounts of short-term assets (relative to long-term assets) and finance the working capital using longer-term sources, such as long-term debt and equity. Benefits of a conservative approach include using more permanent capital with less need for rolling over; greater flexibility during market disruptions; and a high probability of meeting short-term obligations. However, the conservative approach results in higher costs and lower profitability. Additionally, long-term lenders may impose operational constraints such as a minimum interest coverage ratio.

An aggressive approach would be to hold relatively lower levels of short-term assets and finance working capital using short-term debt. The benefit of an aggressive approach is lower costs, but its risks are failing to meet business obligations and vulnerability to market disruptions.

A moderate approach seeks to find a middle ground. Permanent current assets are funded using long-term sources of capital, while variable (seasonal) current assets are funded using short-term sources.

Short-Term Liquidity Sources

Firms should maintain a variety of alternative sources and evaluate the costs of each source, while securing sources ahead of time to meet spikes in liquidity needs.

Factors that affect a firm's approach to short-term funding include the following:

- Company size (smaller firms have limited options)
- Creditworthiness (which affects the interest rate on loans as well as operational restrictions imposed)
- Legal systems (developed economies with well-defined protections for lenders offer more funding alternatives)
- Regulatory concerns (firms in regulated industries, such as banks and utilities, have restrictions on funding sources as well as amounts raised)
- Underlying assets (that serve as collateral on loans)



MODULE QUIZ 23.1

1. Compared to its industry peers, a company with a shorter cash conversion cycle *most likely*:
   A. has more inventory.
   B. has less accounts receivable.
   C. pays its suppliers more promptly.
2. A company receives an invoice of $150,000 for machine tools with terms of "1.5/15 net 40." The cost to the company of delaying payment of this receivable is *most* appropriately described as $2,250 for the use of:
   A. $150,000 for 40 days.
   B. $150,000 for 25 days.
   C. $147,750 for 25 days.
3. Which of the following actions is *most likely* to increase liquidity for a corporation?
   A. Selling inventory at a discount of 5%.
   B. Availing of a discount of 10% by paying accounts payable early.
   C. Extending customers' credit terms from 90 days to 120 days.
4. Which of the following is *least likely* a primary source of liquidity?
   A. Borrowings from banks.
   B. Cash flow from operations.
   C. Delaying capital expenditures.
5. Which of the following *most likely* represents conservative working capital management?
   A. Decreasing inventory on hand to reduce insurance costs.
   B. Financing an increase in receivables by increasing long-term borrowing.
   C. Selling marketable securities and using the proceeds to acquire real estate.

KEY CONCEPTS

LOS 23.a

The cash conversion cycle (CCC) represents the time it takes for a company to convert its investments in inventory and other resources into cash inflows from sales. The CCC increases with an increase in days of inventory on hand (DOH), days sales outstanding (DSO), and decreases with an increase in days payable outstanding (DPO).

$$CCC = DOH + DSO - DPO$$

$$\text{Total working capital} = \text{current assets} - \text{current liabilities}$$

$$\text{Net working capital} = \text{current assets (except cash and marketable securities)} - \text{current liabilities (excluding short-term and current debt)}$$

LOS 23.b

Primary liquidity sources include cash and marketable securities on hand, bank borrowings, and cash generated from the business. Secondary sources of liquidity include forgoing dividends, delaying capital investments, selling assets, issuing equity, restructuring debt, and bankruptcy protection filing. Using secondary liquidity sources sends a negative signal to the market and incurs a cost of liquidity.

A drag on liquidity occurs when cash inflows lag (DOH or DSO increases), while a pull on liquidity occurs when cash outflows accelerate (DPO decreases).

Measures of a company's short-term liquidity include:

- Current ratio = current assets / current liabilities.
- Quick ratio = (cash + marketable securities + receivables) / current liabilities.
- Cash ratio = (cash + marketable securities) / current liabilities.

LOS 23.c

A conservative approach to working capital management involves high levels of current assets financed with long-term debt and equity. Compared to a more aggressive approach, a conservative approach provides more liquidity and involves less financial risk, but has higher financing costs and reduces returns.

An aggressive approach to working capital management involves lower levels of current assets and financing working capital needs with short-term debt. Compared to a more conservative approach, an aggressive approach typically has lower financing costs and results in higher returns, but decreases liquidity and increases financial risk.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 23.1

1. **B** CCC = DOH + DSO − DPO. A company with a shorter CCC would have lower DOH (lower amount of cash tied in inventory), lower DSO (lower amount of cash tied in accounts receivables), or higher DSO (increased use of supplier credit). (LOS 23.a)

2. **C** The terms indicate that the company can pay \$150,000(1 – 0.015) = \$147,750 on day 15 (after the invoice date) or pay \$150,000 on day 40—effectively gaining the use of \$147,750 for 25 days at a cost of \$2,250. (LOS 23.a)

3. **A** Liquidity can be increased by reducing inventory. Extending repayment times would increase DSO, reducing liquidity. Paying suppliers earlier decreases DPO, thus reducing liquidity. (LOS 23.b)

4. **C** Primary sources of liquidity include cash and marketable securities on hand, bank borrowings, and cash generated from the business. Delaying capital expenditure is a secondary source of liquidity. (LOS 23.b)

5. **B** Financing an increase in a current asset with long-term borrowing is an example of conservative working capital management. The other choices describe decreases in current assets and therefore more likely represent aggressive working capital management. (LOS 23.c)

READING 24

CAPITAL INVESTMENTS AND CAPITAL ALLOCATION

MODULE 24.1: CAPITAL INVESTMENTS AND PROJECT MEASURES



Video covering this content is available online.

LOS 24.a: Describe types of capital investments.

The four types of capital investments include going concern projects, regulatory/compliance projects, expansion projects, and other projects.

- **Going concern projects** may be needed to maintain the business or reduce costs. Projects that maintain the business do not require detailed analysis. The only issues are whether the existing operations should continue and, if so, whether existing procedures or processes should be maintained. Projects to improve efficiency may involve determining if equipment that is obsolete, but still usable, should be replaced. To reduce financing risk, companies often use a **match funding** approach, financing projects with capital sources that are consistent with the project life. Analysts often use a company's annual depreciation expense as an estimate of the capital investment it needs for going concern projects.
- **Regulatory/compliance projects** may be required by a government agency or insurance company and often involve safety-related or environmental concerns. These projects typically generate little to no revenue and require the company to evaluate alternative ways of carrying out the projects.
- **Expansion projects** grow the business and require a complex decision-making process that includes forecasting future demand. Expansion projects can involve entering new markets or introducing new products within the same market. A detailed analysis including forecasting revenues and expenses is required.
- Other projects, such as new investments outside a company's existing lines of business, also entail a complex decision-making process with detailed analysis due to the uncertainty involved. Other projects are often similar to startups that explore a new technology or business idea. These projects can also involve buying out an existing company in a new industry, which involve risks that include overpaying.

LOS 24.b: Describe the capital allocation process, calculate net present value (NPV), internal rate of return (IRR), and return on invested capital (ROIC), and

contrast their use in capital allocation.

The **capital allocation process** is identifying and evaluating capital projects (i.e., projects where the cash flows to the firm will be received over a period longer than a year). Any corporate decisions with an impact on future earnings can be examined using this framework. Decisions about buying a new machine, expanding business into another geographic area, moving the corporate headquarters to Cleveland, or replacing a delivery truck, to name a few, can be examined using a capital allocation analysis.

For a number of good reasons, capital allocation may be the most important responsibility that a financial manager has. First, because a capital allocation decision often involves the purchase of costly long-term assets with lives of many years, the decisions made may determine the future success of the firm. Second, the principles underlying the capital allocation process also apply to other corporate decisions, such as working capital management and making strategic mergers and acquisitions. Finally, making good capital allocation decisions is consistent with management's primary goal of maximizing shareholder value.

The capital allocation process has four administrative steps:

Step 1: Idea generation. The most important step in the capital allocation process is generating good project ideas. Ideas can come from a number of sources, including senior management, functional divisions, employees, or sources outside the company.

Step 2: Analyzing project proposals. Because the decision to accept or reject a capital project is based on the project's expected future cash flows, a cash flow forecast must be made for each project to determine its expected profitability.

Step 3: Create the firm-wide capital budget. Firms must prioritize profitable projects according to the timing of the project's cash flows, available company resources, and the company's overall strategic plan. Many projects that are attractive individually may not make sense strategically.

Step 4: Monitoring decisions and conducting a post-audit. It is important to follow up on all capital allocation decisions. An analyst should compare the actual results to the projected results, and project managers should explain why projections did or did not match actual performance. Because the capital allocation process is only as good as the estimates of the inputs into the model used to forecast cash flows, a post-audit should be used to identify systematic errors in the forecasting process and improve company operations.

Net Present Value (NPV)

Net present value (NPV) is the sum of the present values of all the expected incremental cash flows if a project is undertaken. The discount rate used is the firm's cost of capital, adjusted for the risk level of the project. For a normal project, with an initial cash outflow followed by a series of expected after-tax cash inflows, the NPV is the present value of the expected inflows minus the initial cost of the project.

$$NPV = CF_0 + \frac{CF_1}{(1+k)^1} + \frac{CF_2}{(1+k)^2} + \ldots + \frac{CF_n}{(1+k)^n} = \sum_{t=0}^{n} \frac{CF_t}{(1+k)^t}$$

where:

CF_0 = initial investment outlay (a negative cash flow)

CF_t = after-tax cash flow at time *t* (can be positive or negative)

k = required rate of return for project

A positive NPV project is expected to increase shareholder wealth, a negative NPV project is expected to decrease shareholder wealth, and a zero NPV project has no expected effect on shareholder wealth.

For *independent* projects, the *NPV decision rule* is simply to accept any project with a positive NPV and to reject any project with a negative NPV.

EXAMPLE: Net present value

Using the project cash flows presented in Figure 24.1, calculate the NPV of the project and determine whether it should be accepted or rejected. Assume that the cost of capital is 9%.

Figure 24.1: Expected Net After-Tax Cash Flows

Year	Cash Flow
0	–$100
1	25
2	50
3	75

Answer:

$$NPV = -100 + \frac{25}{1.09} + \frac{50}{(1.09)^2} + \frac{75}{(1.09)^3} = 22.93$$

The project has a positive NPV, so it should be accepted.

You may calculate NPV directly by using the cash flow (CF) keys on your calculator. The process is illustrated in Figure 24.2.

Figure 24.2: Calculating NPV With the TI BA II Plus™

Keystrokes	Explanation	Display
[CF] [2nd] [CLR WORK]	Clear memory registers	CF0 = 0.0000
100 [+/−] [ENTER]	Initial cash outlay	CF0 = −100.0000
[↓] 25 [ENTER]	Period 1 cash flow	C01 = 25.0000
[↓]	Frequency of cash flow 1	F01 = 1.0000
[↓] 50 [ENTER]	Period 2 cash flow	C02 = 50.0000
[↓]	Frequency of cash flow 2	F02 = 1.0000
[↓] 75 [ENTER]	Period 3 cash flow	C03 = 75.0000
[↓]	Frequency of cash flow 3	F03 = 1.0000
[NPV] 9 [ENTER]	9% discount rate	I = 9.0000
[↓] [CPT]	Calculate NPV	NPV = 22.9335

Internal Rate of Return (IRR)

For a normal project, the **internal rate of return (IRR)** is the discount rate that makes the present value of the expected incremental after-tax cash inflows just equal to the initial cost of the project. More generally, the IRR is the discount rate that makes the present value of a project's estimated cash inflows equal to the present value of the project's estimated cash outflows. That is, IRR is the discount rate that makes the following relationship hold:

$$\text{PV inflows} = \text{PV outflows}$$

The IRR is also the discount rate for which the NPV of a project is equal to zero:

$$\text{NPV} = 0 = \text{CF}_0 + \frac{\text{CF}_1}{(1+\text{IRR})^1} + \frac{\text{CF}_2}{(1+\text{IRR})^2} + \ldots + \frac{\text{CF}_n}{(1+\text{IRR})^n} = \sum_{t=0}^{n} \frac{\text{CF}_t}{(1+\text{IRR})^t}$$

IRR decision rule: If IRR is greater than the required rate of return, accept the project. If IRR is less than the required rate of return, reject the project.

The required rate of return for a given project is usually the firm's cost of capital. However, a project analyst may adjust the required rate of return above or below the firm's cost of capital to account for differences between the project's risk and the average risk of all of the firm's projects (which is reflected in the firm's current cost of capital).

For this reason, the minimum IRR above which a project will be accepted is often referred to as the **hurdle rate**. Projects with IRRs above this rate will be accepted, while those with IRRs below this rate will not be accepted.

EXAMPLE: Internal rate of return

Continuing with the cash flows presented in Figure 24.1 for the previous example, calculate the IRR of the project and determine whether it should be accepted or rejected. Assume that the required rate of return is 9%.

Answer:

$$0 = -100 + \frac{25}{(1+\text{IRR})} + \frac{50}{(1+\text{IRR})^2} + \frac{75}{(1+\text{IRR})^3}$$

The cash flows should be entered as in Figure 24.2 (if you haven't changed or cleared them, they are still there from the previous calculation of NPV).

With the TI calculator, the IRR can be calculated with [IRR] [CPT] to get 19.4377%.

The project should be accepted because its IRR is greater than the 9% required rate of return.

A project has a **conventional cash flow pattern** if the sign on the cash flows changes only once, with one or more cash outflows followed by one or more cash inflows. An **unconventional cash flow pattern** has more than one sign change. For example, a project might have an initial investment outflow, a series of cash inflows, and a cash outflow for asset retirement costs at the end of the project's life. Another example of unconventional cash flow is when the time intervals between cash flows differ, for example if the first cash flow occurs 1 year after project inception while the second cash flow occurs 1.5 years after the first. Spreadsheet software is most suitable for calculating NPV and IRR with unconventional cash flows.

Relative Advantages and Disadvantages of the NPV and IRR Methods

A key advantage of NPV is that it is a direct measure of the expected increase in the value of the firm. In theory, a positive NPV project should cause a proportionate increase in a company's stock price.

A key advantage of IRR is that it measures profitability as a percentage, showing the return on each dollar invested. The IRR provides information on the margin of safety that the NPV does not. From the IRR, we can tell how much below the IRR (estimated return) the actual project return could fall, in percentage terms, before the project becomes uneconomic (has a negative NPV).

The *disadvantages* of the IRR method are:

- It assumes that project's cash flows are reinvested at the IRR while NPV assumes that those cash flows are reinvested at the project's required rate of return. It is more realistic to assume the latter.
- For multiple sign changes, a project may have multiple IRRs that are difficult to interpret.

Return on Invested Capital

One way to examine whether a company is creating value for its shareholders is to compare the return on the company's investment in assets to its cost of capital. A company's **return on invested capital (ROIC)**, or simply **return on capital**, is defined

as its **net operating profit after tax (NOPAT)**, or simply after-tax operating profit, divided by the average book value of its total capital over the period:

$$\text{return on invested capital} = \frac{\text{net operating profit after tax}}{\text{average book value of invested capital}}$$

After-tax operating profit is net income plus after-tax interest expense. We use this because we want to measure the return to all sources of capital (both debt and equity). Invested capital includes long-term debt and equity, but excludes working capital.

We can rewrite the ROIC equation as:

$$\text{ROIC} = \left(\frac{\text{NOPAT}}{\text{Average Invested Capital}}\right) = \left(\frac{\text{NOPAT}}{\text{Sales}}\right) \times \left(\frac{\text{Sales}}{\text{Average Invested Capital}}\right)$$

We refer to the ratio of after-tax operating profit to sales as *operating margin after tax*, and we refer to the ratio of sales to invested capital as *capital turnover* or *asset turnover*. This means that ROIC is the product of operating margin after-tax and invested capital turnover. A company can increase its ROIC by improving its operating margin or by increasing its capital turnover.

Analysts compare ROIC to investors' required rate of return. In this context, the required rate of return is a blended rate for both debt and equity. If ROIC is greater than the required rate, the firm is adding value over time.

One attraction of ROIC is that it is based on accounting data, which are available to outside investors. While NPV and IRR are project-specific, ROIC is for the firm as a whole. ROIC is more relevant for outside investors because they typically cannot invest in a firm's individual projects.

There are three concerns with using ROIC:

1. Because accounting treatments differ, ROIC may not be comparable across companies.
2. ROIC is backward-looking and can be volatile from year to year.
3. Because ROIC is for the whole company, it may let profitable projects and good decisions mask unprofitable projects and poor decisions.



MODULE QUIZ 24.1

1. Which of the following is *most likely* a going concern project?
   A. Opening a retail outlet in a new region.
   B. Acquiring and merging with a supplier to secure a source for a key component.
   C. Purchasing a new model of a factory machine that will decrease unit production costs.
2. In the capital allocation process, a post-audit is used to:
   A. improve cash flow forecasts and stimulate management to improve operations and bring results into line with forecasts.
   B. improve cash flow forecasts and eliminate potentially profitable but risky projects.
   C. stimulate management to improve operations, bring results into line with forecasts, and eliminate potentially profitable but risky projects.
3. A company is considering the purchase of a copier that costs $5,000. Assume a required rate of return of 10% and the following cash flow schedule:
   - Year 1: $3,000.

- Year 2: $2,000.
- Year 3: $2,000.

The project's NPV is *closest* to:

A. −$309.
B. +$883.
C. +$1,523.

MODULE 24.2: CAPITAL ALLOCATION PRINCIPLES AND REAL OPTIONS



Video covering this content is available online.

LOS 24.c: Describe principles of capital allocation and common capital allocation pitfalls.

Principles of Capital Allocation

The capital allocation process involves the following key principles:

- *Decisions are based on after-tax cash flows, not accounting income*. Accounting income is based on accruals and does not consider the timing of cash flows. The impact of taxes must be considered when analyzing all capital allocation projects. Firm value is based on cash flows firms get to keep, not those they send to the government. Any tax savings from non-cash tax deductions, such as depreciation and amortization, should also be included in the analysis.
- *Incremental cash flows only*. Incremental cash flows are those that change if the project is undertaken. **Sunk costs** are costs that cannot be avoided even if the project is not undertaken. Because these costs are not affected by the accept/reject decision, they should not be included in the analysis. An example of a sunk cost is a consulting fee paid to a marketing research firm to estimate demand for a new product before making a decision on the project. Project analysis should consider the cash flow impact on other parts of the business. A negative effect, called cannibalization, occurs when a new project replaces sales from an existing product. An example of cannibalization is when a soft drink company introduces a diet version of an existing beverage. A positive externality exists when doing the project would have a positive effect on sales of a firm's other product lines.
- *The timing of cash flows is important*. Capital allocation decisions account for the time value of money, which means that cash flows received earlier are worth more than cash flows to be received later.

Common mistakes managers make when evaluating capital projects have two categories: cognitive errors (calculation errors) and behavioral biases (errors of judgment).

Cognitive Errors

- *Poor forecasting.* Examples include incorrectly allocating overhead costs or neglecting to anticipate how competitors will respond to a project.

- *Not considering the cost of internal funds.* Often, the cost of internally generated funds is not accounted for (it should be same as cost of equity because those funds would have been paid as dividends if not for the project). Companies that have an aversion to paying dividends are potentially using the retained earnings to fund poor projects.
- *Incorrectly accounting for inflation*. Firms can analyze projects either in nominal or real terms. An analysis based on real cash flows must also use a real discount rate.

Behavioral Biases

- *Pet projects of senior management*. Projects that have the personal backing of influential members of senior management may contain overly optimistic projections that make the project appear more profitable than it really is. In addition, the project might not be subjected to the same level of scrutiny as other projects.
- *Inertia in setting the entire capital budget*. Capital budgets for many companies are similar from one year to the next. This indicates anchoring of capital budgets to the prior year rather than appropriately considering what opportunities exist every year. In a company with a culture of maximizing shareholder value, managers will return excess funds to shareholders whenever there is a lack of positive NPV projects and make a case for expanding the budget when there are multiple positive NPV opportunities. Analysts should watch for companies with static or rising capital budgets coupled with declining returns as a sign of this bias.
- *Basing investment decisions on EPS or ROE*. Managers whose incentive compensation is tied to increasing EPS or ROE may avoid positive long-term NPV investments that are expected to reduce EPS or ROE in the short run.
- *Failure to generate alternative investment ideas*. Generating investment ideas is a crucial step in the capital allocation process. However, once a manager comes up with a "good" idea, he or she may go with it rather than searching for an idea that is "better."

LOS 24.d: Describe types of real options relevant to capital investments.

Real options are future actions that a firm can take, given that they invest in a project today. Real options are similar to financial options (put and call options) in that they give the option holder the right, but not the obligation, to take a future action. The value of real options could enhance a project's NPV. Options never have negative values because if, in the future, the specified action will have a negative value, the option holder will not take the action (i.e., not exercise the option).

Types of real options include the following:

- **Timing options** allow a company to delay making an investment because it expects to have better information in the future.
- **Abandonment options** allow management to abandon a project if the present value of the incremental cash flows from exiting a project exceeds the present value of the incremental cash flows from continuing the project.

- **Expansion options** or **growth options** allow a company to make additional investments in future projects if the company decides they will create value.
- **Flexibility options** give managers choices regarding the operational aspects of a project. The two main forms are price-setting and production-flexibility options.
  - **Price-setting options** allow the company to change the price of a product. For example, the company may raise prices if demand for a product is high in order to benefit from that demand without increasing production.
  - **Production-flexibility options** may include paying workers overtime, using different materials as inputs, or producing a different variety of product.
- **Fundamental options** are projects that are options themselves because the payoffs depend on the price of an underlying asset. For example, the payoff for a copper mine is dependent on the market price for copper. If copper prices are low, it may not make sense to open a copper mine, but if copper prices are high, opening the copper mine could be very profitable. The operator has the option to close the mine when prices are low and open it when prices are high.

One way companies can include real options in their project analysis is to estimate their value and add it to the NPV (while subtracting any extra cost to acquire the real option). This can incorporate available techniques such as option pricing models or decision trees. Another approach is to simply consider the NPV without real options to be the project's minimum value.

MODULE QUIZ 24.2

1. Which of the following statements concerning the principles underlying the capital allocation process is *most accurate*?
   A. Cash flows should include tax benefits of non-cash expense deductions.
   B. The net income for a project is essential for making a correct capital allocation decision.
   C. Cash flows should be project specific, ignoring sunk costs and impacts on other parts of the business.
2. A manufacturer of clothes washing machines decides to add matching clothes dryers to its product line. In this case, it is *most likely* important in the project analysis to consider:
   A. sunk costs.
   B. negative impact on other parts of the business.
   C. positive impact on other parts of the business.
3. An analyst is estimating the NPV of a project to introduce a new spicier version of its well-known barbeque sauce into its product line. A cost that should *most likely* be excluded from his analysis is:
   A. $200,000 to develop a recipe for the new sauce.
   B. $100,000 for a marketing survey that was conducted to determine demand for a spicier sauce.
   C. a $150,000 decrease in sales of its current sauce as some current customers switch to the spicier sauce.
4. Albert Duffy, a project manager at Crane Plastics, is considering taking on a new capital project. When presenting the project, Duffy shows members of Crane's executive management team that, because the company has the ability to have employees work overtime, the project makes sense. The project Duffy is taking on would be *best* described as having a(n):

A. flexibility option.
B. expansion option.
C. fundamental option.

KEY CONCEPTS

LOS 24.a

Capital investments include going concern projects to maintain a business or to reduce costs, required regulatory/compliance projects, expansion projects, and other projects that increase the size and scope of a company.

LOS 24.b

Capital allocation is the process of evaluating capital projects (i.e., projects with cash flows over a period longer than one year).

Steps of the capital allocation process are: (1) generate investment ideas; (2) analyze project ideas; (3) create a firm-wide capital budget; and (4) monitor decisions and conduct a post-audit.

NPV is the sum of the present values of a project's expected cash flows and represents the change in firm value from undertaking a project. Positive NPV projects should be undertaken, but negative NPV projects should not because they are expected to decrease the value of the firm.

An IRR is the discount rate at which the present values of a project's expected cash inflows and cash outflows are equal (i.e., the discount rate for which the NPV of a project is zero). A project for which the IRR is greater (less) than the appropriate discount rate for the project will have an NPV that is positive (negative) and should be accepted (not accepted).

Return on invested capital can be compared to a company's required rate of return to indicate whether the company has increased or decreased firm value over time.

LOS 24.c

Capital allocation decisions should be based on after-tax cash flows, ignore sunk costs, and capture any spillover effects on other parts of the business. Timing of cash flows is important.

Common mistakes in the capital allocation process include the following:

- Having overly optimistic assumptions for pet projects of senior management
- Basing long-term investment decisions on short-term EPS or ROE considerations
- Poor cash flow estimation, misestimating overhead costs, and improper accounting for inflation
- Not accounting for cost of internally generated funds in the estimate of required rate of return
- Static capital budgets without regard to actual investment opportunities
- Failure to generate alternative investment ideas

LOS 24.d

Real options allow managers to make future decisions that change the value of capital allocation decisions made today.

- Timing options allow a company to delay making an investment.
- Abandonment options allow management to abandon a project if the present value of the incremental cash flows from exiting a project exceeds the present value of the incremental cash flows from continuing a project.
- Expansion options allow a company to make additional investments in a project if doing so creates value.
- Flexibility options give managers choices regarding the operational aspects of a project. The two main forms are price-setting and production-flexibility options.
- Fundamental options are projects that are options themselves because the payoffs depend on the price of an underlying asset.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 24.1

1. **C** Going concern projects are those to maintain the business or to increase the efficiency of existing operations. The other two projects are business growth investments that increase the size of the company. (LOS 24.a)

2. **A** A post-audit identifies what went right and what went wrong. It is used to improve forecasting and operations. (LOS 24.b)

3. **B** $CF_0 = -5{,}000$; $CF_1 = 3{,}000$; $CF_2 = 2{,}000$; $CF_3 = 2{,}000$; I / Y = 10; NPV = \$883. (LOS 24.b)

Module Quiz 24.2

1. **A** Cash flows should be after-tax and include any tax savings from non-cash deductions (e.g., depreciation and amortization). While sunk costs should be ignored, the impact on other parts of the business (positive or negative) should be accounted for by adjusting cash flows. Accounting net income, which includes non-cash expenses, is irrelevant. Incremental cash flows are essential for making correct capital allocation decisions. (LOS 24.c)

2. **C** It is quite possible that offering a matching dryer will increase sales of their washers because some consumers will prefer a matching set. The increased sales of their washers is a positive impact, and those incremental profits should be considered in the analysis. A negative impact on another part of the business would be a consideration if introducing dryers could be expected to decrease washer sales. Sunk costs should not be considered in project analysis. (LOS 24.c)

3. **B** The cost of the marketing survey should not be included because it is a sunk cost; it will be incurred whether they decide to do the project or not. The decrease in

sales of their current sauce if the spicier version is introduced (cannibalization) should be considered in the analysis. The cost of recipe development should be included because it will only be incurred if they decide to go ahead with the introduction of the new spicier sauce. (LOS 24.c)

4. **A** The project described has production flexibility regarding the level of production. Other flexibility options might be to produce a different product or to use different inputs at some future date. Including the value of real options can improve the NPV estimates for individual projects. (LOS 24.d)

READING 25

CAPITAL STRUCTURE

MODULE 25.1: WEIGHTED-AVERAGE COST OF CAPITAL



Video covering this content is available online.

LOS 25.a: Calculate and interpret the weighted-average cost of capital for a company.

An issuer's **weighted-average cost of capital (WACC)** is a blended rate that includes its cost of debt and equity:

$$\text{WACC} = [\text{weight of debt} \times \text{pretax cost of debt} \times (1 - \text{tax rate})] + (\text{weight of equity} \times \text{cost of equity})$$

The cost of debt is lower than the cost of equity because debt has priority of claims over equity. Because most jurisdictions allow tax deductions for corporate interest expense, we adjust the pretax cost of debt to its after-tax value in the WACC formula.

The weights for debt and equity can be target weights or market value weights. Analysts often estimate a company's target weights based on book value of debt and equity. Market value weights are appropriate for estimating the current opportunity cost of capital as they reflect current market conditions.

> **EXAMPLE: WACC**
>
> ABC, Inc.'s, capital structure is 50% debt and 50% equity. ABC's cost of debt is 8%, while the cost of equity is 11%. ABC's corporate tax rate is 30%. Calculate the WACC for ABC.
>
> **Answer:**
>
> $$\text{WACC} = [0.50 \times 0.08 \times (1 - 0.30)] + (0.50 \times 0.11) = 0.083, \text{ or } 8.3\%$$

LOS 25.b: Explain factors affecting capital structure and the weighted-average cost of capital.

Capital structures vary considerably among companies. Typically a company will target a capital structure that minimizes its WACC, while also considering the nature of

its assets (i.e., long-lived or shorter-term) when choosing the duration of its financing sources.

We may view factors that affect capital structures as those that determine a company's capacity to service debt. These factors may be internal or external to a firm. Internal factors include the characteristics of the business or industry, a company's life cycle stage, a company's existing debt level, and the corporate tax rate. External factors include market and business cycle conditions, regulation, and industry norms.

Company characteristics that influence the proportion of debt in a company's capital structure include the following:

- *Growth and stability of revenue.* High growth of revenue or stability of significant revenue suggest a continuing ability to service debt.
- *Growth and predictability of cash flow.* Growing cash flow increases the ability to service debt. Significant and stable cash flows indicate a continuing ability to service debt.
- *Amount of business risk*. More business risk (operational risk and demand risk) means greater variability of earnings and cash flows, which decreases the ability to service debt.
- *Amount and liquidity of company assets.* Assets can be pledged as collateral to make a company's debt more attractive. When assets are more liquid (easier to turn into cash, values more stable), they can be pledged more readily.
- *Cost and availability of debt financing.* Companies find debt relatively more attractive when the cost of debt is lower and investors are more willing to lend to the company.

In general, the more stable, predictable, and recurring are a company's revenues and cash flows, the higher proportion of debt it can have in its capital structure. With this in mind, we can say, other things equal:

- Companies in noncyclical industries are better able to support high proportions of debt than companies in cyclical industries.
- Companies with low fixed operating costs as a proportion of total costs (i.e., low operating leverage) are better able to support high proportions of debt than companies with high fixed costs.
- Companies with subscription-based revenue models are better able to support high proportions of debt than companies with pay-per-use revenue models.

The types of assets companies use to generate revenues also affect their ability to issue and service debt. For example, creditors tend to view tangible assets as better collateral than intangible assets, especially when those assets can be sold for cash, if necessary, without losing significant value (i.e., are more liquid) or are more readily substituted for similar assets (i.e., are more fungible). A company that owns its productive assets outright as opposed to using assets owned by others (such as a franchise model) has more collateral, which improves access to debt financing and reduces borrowing costs.

In addition to the types of assets, other issuer-specific conditions include the levels of existing debt and volatility of revenues and earnings. High levels of debt or high

volatility of earnings indicates a lower ability to issue additional debt. Leverage (e.g., debt-to-equity or debt-to-operating profit) and coverage ratios (e.g., interest coverage = EBIT / interest expense) are used for analyzing debt capacity.

We can see the effects of these various factors on the debt-to-equity ratios of companies at different stages of their company's life cycle. Consider the following three stages:

1. *Start-up stage.* Sales are just beginning, and operating earnings and cash flows tend to be low or negative. Business risk is relatively high. Company debt is quite risky and, if issued, would require high interest rates. Assets, both accounts receivable and fixed assets, typically are low; therefore, they are not available as collateral for debt. For these reasons, start-up companies are financed almost exclusively with equity. In some cases, high-growth companies with rapidly rising stock prices may find it possible to issue **convertible debt**, which allows the company to use a lower-cost source of capital and avoid immediate dilution of existing shareholders' interests. Leasing assets is another source of debt financing available to companies in this stage of the life cycle.
2. *Growth stage.* Revenue and cash flow are rising, and business risk is somewhat reduced. Debt financing cost is somewhat reduced (i.e., usage is conservative), and investors may be willing to lend to the company, often with the loans secured by fixed assets or accounts receivable.
3. *Mature stage.* In this stage, revenue growth is slowing and business risk is much lower. Cash flow is significant and relatively stable. Debt financing, including unsecured debt, is widely available at a relatively low cost.

The cost of capital is also influenced by top-down factors. Top-down factors are macroeconomic factors that affect the benchmark interest rate (e.g., U.S. Treasury rate) and credit spreads to benchmark rates. These include inflation, the real GDP growth rate, monetary policy, and exchange rates. In business cycle downturns, debt investors demand greater yield spreads to benchmark bonds from corporate borrowers to compensate them for the increased risk of default. This is especially true for companies in cyclical industries. Some industries may be favorably affected by the economic environment. For example, oil industry profitability is linked to oil prices, and spreads for issuers in the industry tend to narrow when oil prices rise.



MODULE QUIZ 25.1

1. Alpaca's capital structure is 60% debt with an interest rate of 6%. Alpaca's cost of equity is 12%, and the corporate tax rate is 15%. Alpaca's WACC is *closest* to:
   A. 6.3%.
   B. 7.9%.
   C. 9.7%.

2. A company is *most likely* to be financed only by equity during its:
   A. start-up stage.
   B. growth stage.

C. mature stage.

MODULE 25.2: CAPITAL STRUCTURE THEORIES

Video covering this content is available online.

LOS 25.c: Explain the Modigliani–Miller propositions regarding capital structure.

MM Proposition I (No Taxes): Capital Structure Irrelevance

In 1958, Nobel laureates Franco Modigliani and Merton Miller (we will refer to them as MM) published their seminal work on capital structure theory. In it, MM demonstrate that under certain assumptions, the value of a firm is unaffected by its capital structure. This result is referred to as MM I. The assumptions that lead to MM I are as follows:

- *Capital markets are perfectly competitive*. There are no transactions costs, taxes, or bankruptcy costs.
- *Investors have homogeneous expectations*. They have the same expectations with respect to cash flows generated by the firm.
- *There is riskless borrowing and lending*. Investors can borrow and lend at the risk-free rate.
- *There are no agency costs*. There are no conflicts of interest between managers and shareholders.
- *Investment decisions are unaffected by financing decisions*. Operating income is independent of how the firm is financed.

Intuitively, we can explain MM I in terms of a pie. MM I essentially concludes that the amount of pie available (value of the firm) does not depend on how it is sliced (the capital structure). The value of a firm does not change depending on how the claims to its earnings are divided. This idea is illustrated in Figure 25.1.

Figure 25.1: MM Capital Structure Irrelevance Proposition



Consider why the pie analogy holds. The operating earnings (EBIT) of a firm are available to providers of capital. In a company with no debt, all of the operating earnings are available to equity holders, and the value of the company is the discounted present value of these earnings, with a discount rate that depends on the risk of

(uncertainty about) earnings. If a company is financed partly by debt and partly by equity, operating earnings are divided between debtholders and equity holders. If one entity purchased all the debt and all the equity of the firm, it would be entitled to all of the EBIT of the firm, so the total value of its holdings must equal the value of an all-equity firm—which also has a claim to all the operating earnings of the firm. Under the assumption that operating earnings are unaffected by financing decisions, the total value of debt and equity will be unaffected by the proportions of debt and equity in a firm's capital structure.

MM Proposition II: Cost of Equity and Leverage

MM's second proposition (MM II) is framed in terms of a firm's cost of capital, rather than firm value. Based on the same assumptions as MM I, MM II states that the cost of equity increases linearly as a company increases its proportion of debt financing. The cash flows promised to bondholders have priority, so that equity holders receive the remaining cash flows after the claims of debtholders have been met. Because the cash flows promised to debtholders are more certain (less risky) than the residual cash flows promised to equity holders, the cost of debt will be less than the cost of equity. The greater the amount of debt in a firm's capital structure, the more uncertain are the residual cash flows to equity holders. MM II tells us that as companies increase the proportion of debt financing, the risk of the cash flows to equity holders increases, which increases the cost of equity.

The conclusion of MM II is that the decrease in financing costs from using a larger proportion of (lower-cost) debt is just offset by the increase in the cost of equity, resulting in no change in the firm's WACC.

Given MM II (that the value of the firm, and therefore its WACC, is unaffected by changes in the proportion of debt financing), we can state the relation that must hold between a company's debt-to-equity ratio and its cost of equity:

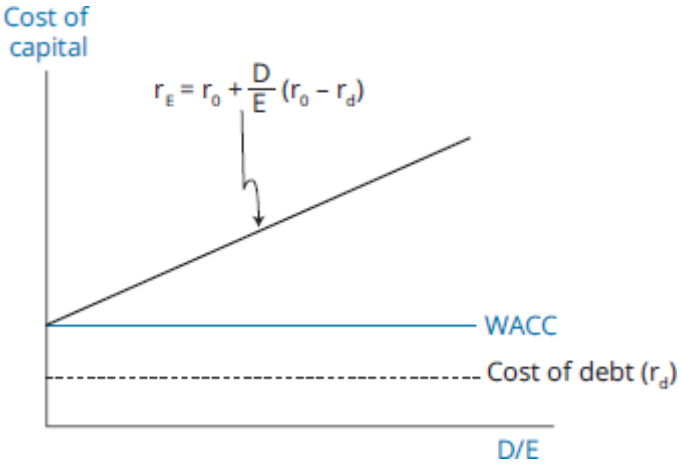
$$r_e = r_0 + \frac{D}{E}(r_0 - r_d)$$

where:

r_e = cost of equity
r_0 = cost of equity with no debt (all equity)
r_d = cost of debt
$\frac{D}{E}$ = debt-to-equity ratio

As leverage (the debt-to-equity ratio) increases, the cost of equity increases, but the cost of debt and WACC are unchanged. This relationship between the cost of equity financing and the debt-to-equity ratio is illustrated in Figure 25.2.

Figure 25.2: MM Proposition II (No Taxes)



MM II is consistent with MM I; if the benefits of greater use of lower-cost debt financing are just offset by the increased cost of equity, the proportions of debt versus equity in the firm's capital structure do not affect the firm's overall cost of capital or the value of the firm.

MM With Taxes: Value Is Maximized at 100% Debt

As is often done in theoretical work, MM started with several simplifying assumptions. The next step is to examine the effects of relaxing some of those assumptions. Here, we examine the MM propositions under the assumptions that earnings are taxed and that interest payments to debtholders are tax deductible. Under the tax code of most countries, interest payments are a pretax expense and are therefore tax deductible, while dividends paid to equity holders are not tax deductible.

This differential tax treatment encourages firms to use debt financing because debt provides a **tax shield** that adds value to the company. The tax shield is equal to the tax rate multiplied by the amount of debt in the capital structure. In other words, the value of a levered firm is equal to the value of an unlevered firm plus the value of the tax shield provided by debt financing.

To continue our analogy of a pie, with the introduction of taxes, the government gets a slice of the pie. When debt financing is used, the government's slice of the pie is smaller, so that the amount of pie available to debt and equity holders is greater. The use of debt reduces taxes due to deductibility of interest expense giving rise to debt tax shield. If a firm were 100% financed with debt, the taxes avoided would be at a maximum, and the after-tax cash flows of the firm (size of the pie) would be maximized. The conclusion of *MM I with taxes* is that the value of the firm is maximized with 100% debt financing.

If we assume a positive tax rate, the formula to solve for return on equity can be used to illustrate *MM II with taxes*:

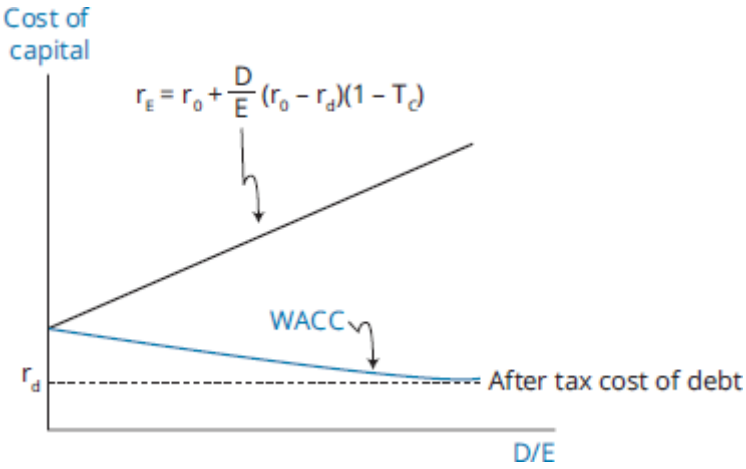
$$r_e = r_0 + \frac{D}{E}(r_0 - r_d)(1 - T_C)$$

where:

T_C = tax rate

Figure 25.3 illustrates that the tax shield provided by debt causes the WACC to decline as leverage increases. The value of the firm is maximized at the point where the WACC is minimized, which is 100% debt.

Figure 25.3: MM Proposition II (With Taxes)



In practice, we do not find that companies finance their assets with 100% debt, or close to it. MM suggested that differential investor tax rates on dividends and interest income could explain differences in capital structures.

Current theory suggests that differences in value-maximizing choices of how much financial leverage to use are the result of additional costs of using debt financing that we have not yet considered—we assumed that there are no bankruptcy costs. However, we can expect these costs, also called the **costs of financial distress**, to increase at higher levels of debt financing. Costs of financial distress are the increased costs a company faces when earnings decline to the point where the firm has trouble paying its fixed financing costs (interest on debt). The expected costs of financial distress for a firm have two components:

1. *Costs of financial distress and bankruptcy* can be direct or indirect. Direct costs of financial distress include the cash expenses associated with the bankruptcy, such as legal fees and administrative fees. Indirect costs include foregone investment opportunities and the costs that result from losing the trust of customers, creditors, suppliers, and employees. Additionally, during periods of financial distress, conflicts of interest between managers (who represent equity owners) and debtholders impose additional costs, referred to as the **agency costs of debt**.
2. *The probability of financial distress* is related to the firm's use of operating and financial leverage. In general, higher amounts of financial leverage increase the probability of financial distress (higher probability that cash flows will fall to an amount that is insufficient to make their promised debt payments). Other factors to consider include the quality of a firm's management and the company's corporate governance structure. Lower-quality management and poor corporate governance lead to a higher probability of financial distress.

Higher expected costs of financial distress tend to discourage companies from using large proportions of debt in their capital structures, all else equal. The expected costs of financial distress shrink our pie as the proportion of debt financing is increased.

LOS 25.d: Describe optimal and target capital structures.

Static Tradeoff Theory

The **static tradeoff theory** seeks to balance the costs of financial distress with the tax shield benefits from using debt. There is an amount of debt financing at which the increase in the value of the tax shield from additional borrowing is exceeded by the value reduction of higher expected costs of financial distress. This point represents the **optimal capital structure** for a firm, where the WACC is minimized and the value of the firm is maximized.

Accounting for the costs of financial distress, the expression for the value of a levered firm becomes the following:

$$V_L = V_U + (t \times \text{debt}) - \text{PV (costs of financial distress)}$$

We illustrate this relationship in Figure 25.4.

Figure 25.4: Static Tradeoff Theory: Cost of Capital vs. Capital Structure

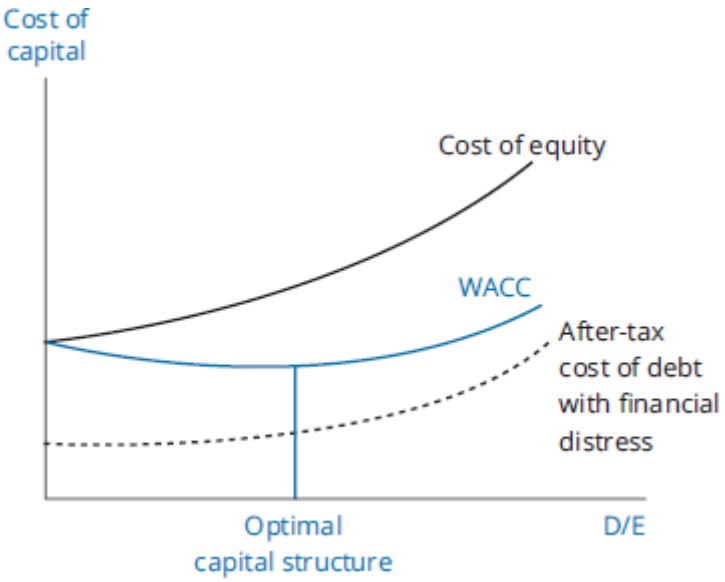
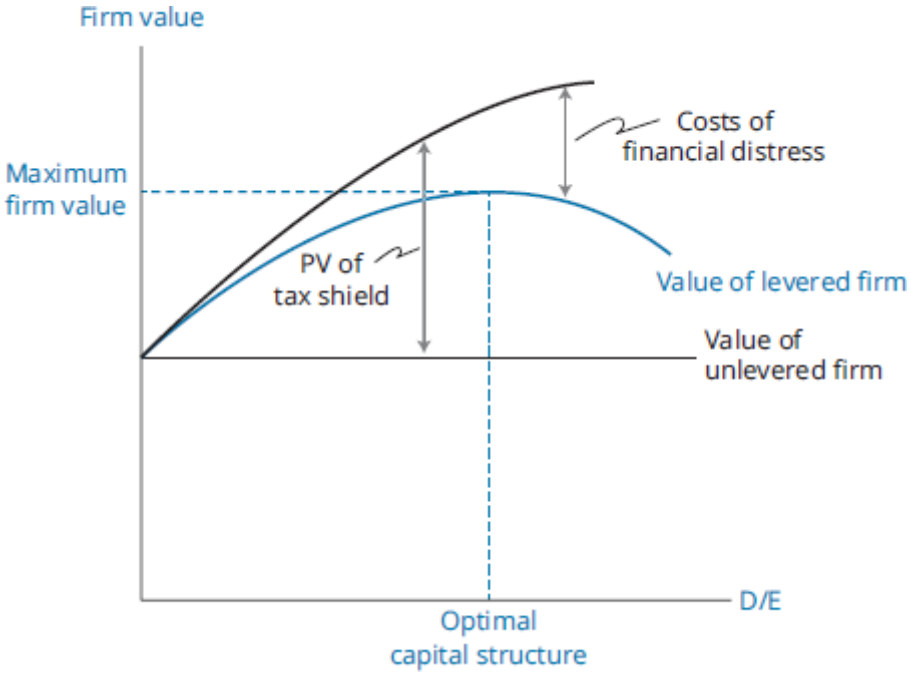


Figure 25.5 illustrates the tradeoff between firm value and capital structure. Note that each firm's optimal capital structure depends on its business risk (operating risk and sales risk), tax rate, corporate governance, industry influences, and other factors.

Figure 25.5: Static Tradeoff Theory: Firm Value vs. Capital Structure



To summarize:

- *MM's propositions with no taxes* or costs of financial distress are that a company's capital structure is irrelevant because its WACC and its value (the discounted present value of its operating earnings) are unchanged by changes in capital structure.
- *MM's propositions with taxes but without costs of financial distress* are that a company's WACC is minimized and its value is maximized with 100% debt.
- *Static tradeoff theory* indicates that firm value initially increases (and WACC decreases) with additional debt financing, but company value decreases at some point when the increase in the expected value of financial distress outweighs the tax benefits of additional debt.

Target capital structure is the capital structure that a firm seeks to achieve on average over time to maximize firm value. That is, it reflects management's beliefs about its optimal capital structure as well as other internal and external factors.

For analysis, the weights to use when estimating a firm's WACC should be based on its target capital structure, the proportions (based on market values) of debt, preferred stock, and equity that the firm expects to achieve over time. An analyst may use management's stated target proportions of equity and debt; however, most firms do not provide these.

In the absence of stated capital structure weights, an external analyst must estimate a firm's target capital structure. Alternatives for estimating target capital structure include the following:

- An analyst may simply use the firm's current capital structure (based on market values) as the best indication of its target capital structure.

- If there has been a noticeable trend in the firm's capital structure, an analyst may incorporate this trend into her estimate of the firm's target.
- An analyst may use the average capital structure weights for a firm's industry.

In contrast with independent analysts, company managers often focus on book values of debt and equity. Reasons book values may be appropriate for internal analysis include:

- Short-term fluctuations in the market values of debt and equity do not really affect a company's appropriate level of debt.
- While market values reflect the perspectives of investors in a company, management's perspective has more to do with how the company needs to deploy its capital to projects.
- Credit rating agencies use book values in their measures of credit quality. Managers may make capital structure decisions with these measures in mind.

In practice, a firm's actual capital structure tends to fluctuate around the target capital structure. Market fluctuations, especially in the market value of a firm's equity, may cause the firm's capital structure to vary from the target. Management may choose to exploit opportunities in a specific financing source. For example, a temporary rise in the firm's stock price may create a good opportunity to issue additional equity, which would result in a higher percentage of equity than the target. Also, external capital is raised in minimum-size lots; hence, adherence to precise weights may not be possible.

Costs of asymmetric information arise from the fact that managers typically have more information about a company's prospects and future performance than owners or creditors. Firms with complex products or little transparency in financial statements tend to have higher costs of asymmetric information, which result in higher required returns on both debt and equity capital.

Because shareholders and creditors are aware that asymmetric information problems exist, these investors look for management behavior that signals what knowledge or opinions management may have about the firm's prospects. For example, taking on the commitment to make fixed interest payments through debt financing sends a signal that management is confident in the firm's ability to make these payments in the future. By contrast, issuing equity is typically viewed as a negative signal that managers believe a firm's stock is overvalued. The cost of asymmetric information increases with the proportion of equity in the capital structure.

Agency costs of equity are related to conflicts of interest between managers and owners. Managers who do not have a stake in the company do not bear the costs associated with excessive compensation or taking on too much (or too little) risk. Because shareholders are aware of this conflict, they take steps to reduce these costs. The result is called the **net agency cost of equity**. Net agency costs of equity have three components:

1. *Monitoring costs* are associated with supervising management and include the expenses of reporting to shareholders and paying the board of directors. Strong corporate governance systems reduce monitoring costs.

2. *Bonding costs* relate to assuring shareholders that the managers are working in the shareholders' best interest. Examples of bonding costs include premiums for insurance to guarantee performance and implicit costs associated with noncompete agreements.
3. *Residual losses* may occur even with adequate monitoring and bonding provisions because such provisions do not provide a perfect guarantee.

According to the **free cash flow hypothesis**, the use of debt forces managers to be disciplined with regard to how they spend cash because they have less free cash flow to use for their own benefit. It follows that greater amounts of financial leverage tend to reduce agency costs.

Pecking order theory, based on asymmetric information, is related to the signals that management sends to investors through its financing choices. According to pecking order theory, managers prefer to make financing choices that are least likely to send negative signals to investors. Financing choices under pecking order theory follow a hierarchy based on visibility to investors. Internally generated capital is most preferred, debt is the next-best choice, and external equity is the least preferred financing option. Pecking order theory implies that the capital structure is a by-product of individual financing decisions.



MODULE QUIZ 25.2

1. A company's optimal capital structure:
   A. maximizes firm value and minimizes the weighted-average cost of capital.
   B. minimizes the interest rate on debt and maximizes expected earnings per share.
   C. maximizes expected earnings per share and maximizes the price per share of common stock.
2. Which of the following statements regarding Modigliani and Miller's Proposition II with no taxes is *most accurate*?
   A. A firm's cost of debt financing increases as a firm's financial leverage increases.
   B. A firm's weighted-average cost of capital is not affected by its choice of capital structure.
   C. A firm's cost of equity financing increases as the proportion equity in a firm's capital structure is increased.
3. To determine their target capital structures in practice, it is *least likely* that firms will:
   A. use the book value of their debt to make financing decisions.
   B. match the maturities of their debt issues to specific firm investments.
   C. determine an optimal capital structure based on the expected costs of financial distress.
4. The pecking order theory of financial structure decisions:
   A. is based on information asymmetry.
   B. suggests that debt is the first choice for financing an investment of significant size.
   C. suggests that debt is the riskiest and least preferred source of financing.

KEY CONCEPTS

LOS 25.a

WACC = [weight of debt × pretax cost of debt × (1 − tax rate)] + (weight of equity × cost of equity)

LOS 25.b

Capital structures vary among companies. Internal factors that affect capital structures include the characteristics of the business, the company's existing debt level, their corporate tax rate, and the company's life cycle stage. External factors include market and business cycle conditions.

A company's ability to issue debt is greater with predictable cash flows sufficient to make required debt payments, and with liquid tangible assets that the company can pledge as collateral for debt.

New companies with few assets and negative or uncertain cash flows will use little to no debt. Growth companies with positive cash flows and decreasing business risk may use lower levels of debt. Mature companies with predictable cash flows tend to use significantly more debt.

LOS 25.c

MM's propositions with no taxes are that a company's capital structure is irrelevant, because its WACC and firm value (the discounted present value of its operating earnings) are unchanged by changes in capital structure.

MM's propositions with taxes are that a company's WACC is minimized and its value is maximized with 100% debt financing.

LOS 25.d

Static tradeoff theory adds the expected costs of financial distress to the model. It indicates that firm value initially increases (and WACC decreases) with additional debt financing, but that company value decreases at some point with additional debt as the increase in the expected costs of financial distress outweigh the increase in tax benefits from additional debt.

In practice, a company's capital structure will fluctuate around the target due to management's exploitation of market opportunities and market value fluctuations (especially of equity) over time.

Pecking order theory is based on information asymmetry between firm management and investors and suggests that management's choice of financing method signals their beliefs about firm value. The theory concludes that retained earnings are the most preferred source of funds, followed by debt financing, and then issuing new equity.

Under the free cash flow hypothesis, the agency costs of equity, which arise because management and shareholders may have conflicting interests, are reduced by increased debt issuance.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 25.1

1. **B** Weight of equity = 1 − 0.60 = 0.40, or 40%. WACC = [0.60 × 0.06 × (1 − 0.15)] + (0.4 × 0.12) = 7.86.

   (LOS 25.a)

2. **A** During the start-up stage, a firm is unlikely to have positive earnings and cash flows or significant assets that can be pledged as debt collateral, so firms in this stage are typically financed by equity only. (LOS 25.b)

Module Quiz 25.2

1. **A** The optimal capital structure minimizes the firm's WACC and maximizes the firm's value (stock price). (LOS 25.c)

2. **B** MM's Proposition II (with no taxes) states that capital structure is irrelevant because the decrease in a firm's WACC from additional debt financing is just offset by the increase in WACC from a decrease in equity financing. The cost of debt is held constant, and the cost of equity financing increases as the proportion of *debt* in the capital structure is increased. (LOS 25.c)

3. **C** While it is a useful theoretical concept, in practice, determining an optimal capital structure based on the cost savings of debt and the expected costs of financial distress is not feasible. Because debt rating companies often use book values of debt, firms often use book values of debt when choosing financing sources. It is common for firms to match debt maturities to the economic lives of specific investments. (LOS 25.d)

4. **A** Pecking order theory is based on information asymmetry and the resulting signals that different financing choices send to investors. It suggests that retained earnings are the first choice for financing an investment, and issuing new equity is the least preferred choice. (LOS 25.d)

READING 26

BUSINESS MODELS

MODULE 26.1: BUSINESS MODEL FEATURES AND TYPES



Video covering this content is available online.

LOS 26.a: Describe key features of business models.

A successful firm must provide a product or service, find customers, deliver the product or service, and make a profit. A **business model** explains how a firm either does or proposes to do this.

"How we will provide it, sell it, and make a profit" is clearly an oversimplification, but this is the essence of a business model. In practice, the answers to these questions have many facets. Here, we present a framework that incorporates some of the complexities involved. It involves answering the questions of *who, what, how, where,* and *how much.*

A business model should do the following:

- *Identify the firm's potential customers (who),* how they are acquired, the cost of customer acquisition, and how the company will monitor and maintain customer satisfaction. Potential customers can be defined in innumerable ways, ranging from every consumer within a geographic area, to dog owners, to only the home country's military in the case of a weapon. This characterization of customer groups is called **segmenting**.
- *Describe the key assets and suppliers of the firm (how).* Key assets may be, for example, a patent, software, or skilled employees. Key suppliers may be a battery manufacturer for an electric vehicle company, a lithium miner for a battery maker, or a supplier of large excavation machinery for a lithium miner.
- *Describe the firm's product or service (what),* how it meets a need for its potential customers, and what differentiates its products from those of competitors (e.g., low price, premium quality, innovative features).
- *Explain how the firm will sell its product or service (where)*, such as online, physical location, direct mail, trade shows, or through sales representatives; whether the firm will sell directly to the buyers (**direct sales**) or use intermediaries such as wholesalers, retailers, agents, or franchisees; and how the firm will deliver its product or service. The answers to these questions comprise a firm's **channel strategy**. A strategy that includes both digital and physical channels, such as internet sales with delivery at a physical location, is referred to as an **omnichannel strategy**.

Firms that sell to other businesses are said to be **B2B** (business to business) firms, while firms that sell to consumers are said to be **B2C** (business to consumer) firms.

- *Explain its pricing strategy (how much)* and why buyers will pay that price for its product, given the competitive landscape of the market.

Pricing Strategies

Companies with undifferentiated products, called **commodity producers**, are price takers (e.g., oil and gas, home loans). Industries with few competitors or highly differentiated products enjoy pricing power (e.g., patented drugs). Companies may be able to charge a premium price by including services or features that are valued by customers.

Price discrimination refers to setting different prices for different customers or identifiable groups of customers. Common examples are **tiered pricing** (based on volume of purchases); **dynamic pricing** (depending on the time of day or day of the week), such as peak and off-peak pricing and low-priced airline tickets for very early or very late flights; **value-based pricing** (e.g., a new drug that offers a marginally better protection compared to existing drugs); and **auction pricing** (e.g., eBay).

Pricing models for multiple products include the following:

- **Bundling.** Where multiple products are complimentary (e.g., a furnished apartment), bundling the products may be a profitable strategy.
- **Razors-and-blades.** A company may find it profitable to sell a piece of equipment for a relatively low price (low margins) and make profits by selling a consumable used with the equipment. Printers and ink cartridges, and an e-reader and e-books are common examples.
- **Add-on pricing.** Options or add-ons priced with high margins are added to the product after the purchase decision has been made. An example is the many pricey options that may be offered after a customer has decided to purchase an automobile.

Other pricing models include the following:

- **Penetration pricing.** A company offers a product at low margins or even at a loss for a period of time to grow market share and achieve greater scale of operations. Netflix followed this strategy to grow its subscriber base rapidly.
- **Freemium pricing.** Offer a product with basic functionality at no cost, but sell or unlock other functionality for a fee. Video game makers have used this strategy to encourage wide usage and then profit on sales of greater functionality (e.g., upgraded weapons).
- **Hidden revenue.** Online content may be "free" but generate revenue through ads. For example, an internet search is free to the user while the search engine's revenue comes from selling user data.

Models that offer alternatives to outright purchases include the following:

- **Subscription model.** Microsoft's model for software has changed from selling the software to a subscription (paying monthly for access) to their Office suite of

software.

- **Licensing and franchising.** For a biotech company that has developed a new and effective drug, it may be most profitable to license the production of the drug to an established drug maker with a large sales force and established distribution channels, rather than developing those resources itself for the single drug. A franchisee typically is permitted to sell in a specific area and pays a percentage of sales to the franchisor, which provides some level of product and marketing support.

Value Proposition and Value Chain

A firm's **value proposition** refers to how customers will value the characteristics of the product or service, given the competing products and their prices. How the firm executes its value proposition is referred to as its **value chain**. A firm's value chain comprises the assets of the firm and how its organization will add value and exploit the firm's competitive advantage. The value chain encompasses the quality of the product, its functionality, service that is included, the sale process, and pricing relative to the competition.

In his 1985 book *Competitive Advantage,* Michael Porter presents five activities (value chain analysis) in which firms should strive to execute well:

1. Inbound logistics
2. Operations
3. Outbound logistics
4. Marketing
5. Sales and service

LOS 26.b: Describe various types of business models.

Conventional business models tend to be industry-specific. Examples include those followed by natural resource producers, manufacturers, distributors, retailers, banks, brokers, service providers, and software. Other business models include the following:

- *Private label manufacturers or contract manufacturers*. Companies produce products for others to market under their own brand name (e.g., Costco's Kirkland-branded products).
- *Licensing agreements*. A company brand is used by another company on its products for a fee, such as a lunch box branded with a Marvel character.
- *Value-added resellers*. They offer such things as installation, service, support, or customization for complex equipment.

Innovations in business models often occur due to new entrants sometimes bringing new technology into the industry to challenge the incumbents. Examples include pricing of software as a service (SaaS), ultra-low-cost airlines, and discount brokers.

Network effects refer to the increase in the value of a network as its user base grows. Examples of this include WhatsApp, eBay, and Facebook. Network effects support an initial strategy of penetration pricing. Network effects capitalize on both size and

scope. Some networks are two-sided or multi-sided—such as Airbnb, which has a multitude of hosts and guests.

Crowdsourcing models benefit from user contributions—content in the case of Wikipedia, traffic conditions and events in the case of Waze, and product improvements or new applications in the case of open-source software.



MODULE QUIZ 26.1

1. A business model is *least likely* to include details about a company's:
   A. largest customers.
   B. workforce characteristics.
   C. revenue and expense estimates.
2. A pricing strategy of offering a basic product at no cost, and unlocking other functionality for a fee, is *most accurately* called:
   A. optional products.
   B. freemium pricing.
   C. subscription pricing.
3. A start-up business provides consumers with professional chefs at home. For its business model to work, the company needs a sufficient number of chefs available to their customers when they have a need. Similarly, sufficient demand from customers is needed to keep the chefs interested in making themselves available. Which of the following is the company's business model *most likely* reliant on?
   A. Licensing.
   B. Bundling.
   C. Network effects.

KEY CONCEPTS

LOS 26.a

A business model should identify a firm's potential customers, describe its products or services and explain how it will sell them, describe its key assets and suppliers, and explain its pricing strategy.

A value proposition refers to how a firm's customers will value the characteristics of the product or service. A value chain refers to how a firm executes its value proposition.

LOS 26.b

Apart from industry-specific conventional business models, other business models include private label manufacturers, value-added resellers, and licensing agreements.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 26.1

1. **C** Detailed forecasts of revenue in expenses would be in a financial plan, but typically not in a business model. A firm's largest customers and information

about its workforce and its value are likely elements of a business model. (LOS 26.a)

2. **B** Freemium pricing offers a product with basic functionality at no cost, but sells or unlocks other functionality for a fee. Optional products involve selling additional products after a purchase decision is made. (LOS 26.a)

3. **C** Network effects refer to the increase in the value of the business as its user base grows. In this instance, the network is two-sided; a sufficient number of customers and service providers are needed for the business to succeed. (LOS 26.b)

TOPIC QUIZ: CORPORATE ISSUERS

You have now finished the Corporate Issuers topic section. Please log into your Schweser online dashboard and take the Topic Quiz on this section. The Topic Quiz provides immediate feedback on how effective your study has been for this material. Questions are more exam-like than typical Module Quiz or QBank questions; a score of less than 70% indicates that your study likely needs improvement. These tests are best taken timed; allow 1.5 minutes per question.

FORMULAS

Quantitative Methods

$$FV = PV\,(1 + r)^t$$

$$PV = \frac{FV}{(1 + r)^t} = FV(1 + r)^{-t}$$

where:

r = interest rate per compounding period
t = number of compounding periods

continuous compounding:

$$FV = PV \times e^{rt}$$

$$PV = FV \times e^{-rt}$$

$$\text{PV of a perpetuity} = \frac{\text{payment}}{r}$$

$$\text{annuity payment} = \frac{r \times PV}{1 - (1 + r)^{-t}}$$

where:

r = interest rate per period
t = number of periods
PV = present value (principal)

$$\text{constant growth dividend discount model: } V_0 = \frac{D_1}{k_e - g_c}$$

$$\text{required rate of return: } k_e = \frac{D_1}{V_0} + g_c$$

$$\text{implied growth rate: } g_c = k_e - \frac{D_1}{V_0}$$

$$\text{sample mean: } \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\text{range} = \text{maximum value} - \text{minimum value}$$

$$\text{mean absolute deviation} = \frac{\sum_{i=1}^{n} |X_i - \overline{X}|}{n}$$

$$\text{sample variance: } s^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n - 1}$$

$$\text{sample standard deviation: } s = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n - 1}}$$

$$\text{coefficient of variation} = \frac{s_x}{\overline{X}} = \frac{\text{standard deviation of x}}{\text{average value of x}}$$

target downside deviation:

$$s_{target} = \sqrt{\frac{\sum_{\text{all } X_i < B}^{n} (X_i - B)^2}{n - 1}}$$

where B = target value

sample covariance:

$$s_{X,Y} = \frac{\sum_{i=1}^{n} \left\{ \left[X_i - \overline{X} \right] \left[Y_i - \overline{Y} \right] \right\}}{n - 1}$$

where:

X_i = an observation of variable X
Y_i = an observation of variable Y
$\overline{X}$ = mean of variable X
$\overline{Y}$ = mean of variable Y
n = number of periods

correlation coefficient: $\rho_{XY} = \frac{s_{XY}}{s_X s_Y}$

Bayes' formula: $P(I \mid O) = \frac{P(O \mid I)}{P(O)} \times P(I)$

slope coefficient for a simple linear regression: $\hat{b}_1 = \frac{cov_{XY}}{\sigma_X^2}$

intercept term for a simple linear regression:

$$\hat{b}_0 = \overline{Y} - \hat{b}_1 \overline{X}$$

where:
$\overline{Y}$ = mean of Y
$\overline{X}$ = mean of X

total sum of squares (SST): $SST = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$

sum of squares regression (SSR): $SSR = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2$

sum of squared errors (SSE): $SSE = \sum_{i=1}^{n} (Y_i - \hat{Y})^2$

coefficient of determination: $R^2 = SSR / SST$

F-statistic:

$$F = \frac{MSR}{MSE} = \frac{SSR/k}{SSE/n - k - 1}$$

where:

MSR = mean regression sum of squares
MSE = mean squared error

Economics

$$\text{real P/B exchange rate} = \text{nominal P/B exchange rate} \times \left(\frac{\text{CPI}_{\text{base currency}}}{\text{CPI}_{\text{price currency}}}\right)$$

$$\text{arbitrage-free forward exchange rate relationship: } \frac{\text{forward}_{d/f}}{\text{spot}_{d/f}} = \frac{(1 + r_{\text{domestic}})}{(1 + r_{\text{foreign}})}$$

Corporate Issuers

$$\begin{aligned}\text{cash conversion cycle} = {} & \text{days of inventory on hand} + \text{days sales outstanding} \\ & - \text{days payables outstanding}\end{aligned}$$

$$\text{Cost of supplier financing} = \left(1 + \frac{a}{1-a}\right)^{\left(\frac{365}{c-b}\right)} - 1$$

where:
a = percent discount
b = days until discount expires
c = days until full payment is due

$$\text{total working capital} = \text{current assets} - \text{current liabilities}$$

$$\begin{aligned}\text{net working capital} = {} & \text{current assets (except cash and marketable securities)} \\ & - \text{current liabilities (excluding short-term and current debt)}\end{aligned}$$

$$\text{current ratio} = \frac{\text{current assets}}{\text{current liabilities}}$$

$$\text{quick ratio} = \frac{(\text{cash and marketable securities} + \text{accounts receivable})}{\text{current liabilities}}$$

$$\text{cash ratio} = \frac{\text{cash and marketable securities}}{\text{current liabilities}}$$

Net present value (NPV):

$$\text{NPV} = \text{CF}_0 + \frac{\text{CF}_1}{(1+k)^1} + \frac{\text{CF}_2}{(1+k)^2} + \ldots + \frac{\text{CF}_n}{(1+k)^n} = \sum_{t=0}^{n} \frac{\text{CF}_t}{(1+k)^t}$$

where:

CF_0 = initial investment outlay

CF_t = after-tax cash flow at time t

k = required rate of return

$$\text{return on invested capital} = \frac{\text{net operating profit after tax}}{\text{average book value of total capital}}$$

APPENDICES

APPENDIX A: AREAS UNDER THE NORMAL CURVE

Most of the examples in this book have used one version of the *z*-table to find the area under the normal curve. This table provides the cumulative probabilities (or the area under the entire curve to left of the z-value).

Probability Example

Assume that the annual earnings per share (EPS) for a large sample of firms is normally distributed with a mean of $5.00 and a standard deviation of $1.50. What is the approximate probability of an observed EPS value falling between $3.00 and $7.25?

If $EPS = x = \$7.25$, then $z = (x - \mu)/\sigma = (\$7.25 - \$5.00)/\$1.50 = +1.50$

If $EPS = x = \$3.00$, then $z = (x - \mu)/\sigma = (\$3.00 - \$5.00)/\$1.50 = -1.33$

Solving Using the Cumulative *Z*-Table

For z-value of 1.50: Use the row headed 1.5 and the column headed 0 to find the value 0.9332. This represents the area under the curve to the left of the critical value 1.50.

For z-value of −1.33: Use the row headed 1.3 and the column headed 3 to find the value 0.9082. This represents the area under the curve to the left of the critical value +1.33. The area to the left of −1.33 is 1 − 0.9082 = 0.0918.

The area between these critical values is 0.9332 − 0.0918 = 0.8414, or 84.14%.

Hypothesis Testing—One-Tailed Test Example

A sample of a stock's returns on 36 nonconsecutive days results in a mean return of 2.0%. Assume the population standard deviation is 20.0%. Can we say with 95% confidence that the mean return is greater than 0%?

$H_0: \mu \leq 0.0\%$, $H_a: \mu > 0.0\%$. The test statistic = *z*-statistic

$$= \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = (2.0 - 0.0) / (20.0 / 6) = 0.60$$

The significance level = 1.0 − 0.95 = 0.05, or 5%. Because we are interested in a return greater than 0.0%, this is a one-tailed test.

Using the Cumulative *Z*-Table

Because this is a one-tailed test with an alpha of 0.05, we need to find the value 0.95 in the cumulative *z*-table. The closest value is 0.9505, with a corresponding critical *z*-value of 1.65. Because the test statistic is less than the critical value, we fail to reject H_0.

Hypothesis Testing—Two-Tailed Test Example

Using the same assumptions as before, suppose that the analyst now wants to determine if he can say with 99% confidence that the stock's return is not equal to 0.0%.

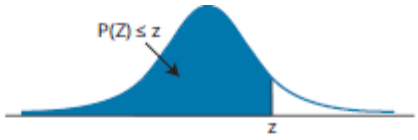
H_0: μ = 0.0%, H_a: μ ≠ 0.0%. The test statistic (*z*-value) = (2.0 − 0.0) / (20.0 / 6) = 0.60. The significance level = 1.0 − 0.99 = 0.01, or 1%. Because we are interested in whether or not the stock return is nonzero, this is a two-tailed test.

Using the Cumulative *Z*-Table

Because this is a two-tailed test with an alpha of 0.01, there is a 0.005 rejection region in both tails. Thus, we need to find the value 0.995 (1.0 − 0.005) in the table. The closest value is 0.9951, which corresponds to a critical *z*-value of 2.58. Because the test statistic is less than the critical value, we fail to reject H_0 and conclude that the stock's return equals 0.0%.

CUMULATIVE *Z*-TABLE

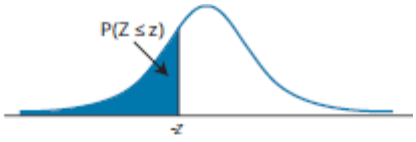
Standard Normal Distribution



$P(Z \leq z) = N(z)$ for $z \geq 0$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Standard Normal Distribution



$P(Z \leq z) = N(z)$ for $z \geq 0$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
–0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
–0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
–0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
–0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
–0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
–0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
–0.7	0.2420	0.2389	0.2358	0.2327	0.2297	0.2266	0.2236	0.2207	0.2177	0.2148
–0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
–0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
–1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
–1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
–1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1057	0.1038	0.1020	0.1003	0.0985
–1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
–1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
–1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
–1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
–1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
–1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
–1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
–2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
–2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
–2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
–2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
–2.4	0.0082	0.0080	0.0078	0.0076	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
–2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
–2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
–2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
–2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
–2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
–3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010

APPENDIX B: STUDENT'S *t*-DISTRIBUTION

Level of Significance for One-Tailed Test						
df	**0.100**	**0.050**	**0.025**	**0.01**	**0.005**	**0.0005**

Level of Significance for Two-Tailed Test						
df	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.599
3	1.638	2.353	3.182	4.541	5.841	12.294
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.768
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

APPENDIX C: *F*-TABLE AT 5% (UPPER TAIL)

F-Table, Critical Values, 5% in Upper Tail

Degrees of freedom for the numerator along top row

Degrees of freedom for the denominator along side row

	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	6.01	2.94	2.90	2.86	2.83
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39

APPENDIX D: *F*-TABLE AT 2.5% (UPPER TAIL)

F-Table, Critical Values, 2.5% in Upper Tails

Degrees of freedom for the numerator along top row

Degrees of freedom for the denominator along side row

	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40
1	648	799	864	900	922	937	948	957	963	969	977	985	993	997	1001	1006
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.41	4.36	4.31
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48

APPENDIX E: CHI-SQUARED TABLE

Values of χ^2 (Degrees of Freedom, Level of Significance)

Probability in Right Tail

Degrees of Freedom	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
1	0.000157	0.000982	0.003932	0.0158	2.706	3.841	5.024	6.635	7.879
2	0.020100	0.050636	0.102586	0.2107	4.605	5.991	7.378	9.210	10.597
3	0.1148	0.2158	0.3518	0.5844	6.251	7.815	9.348	11.345	12.838
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	4.107	5.009	5.892	7.041	19.812	22.362	24.736	27.688	29.819
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	12.878	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.994
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.335
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
50	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
80	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
100	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.170

INDEX

A

abandonment options, 283
accredited investors, 242
action lag, 179
activist shareholder, 258
additivity principle, 27
ad hoc committee, 258
after-tax nominal return, 12
agency costs, 255
agency costs of debt, 296
agency costs of equity, 299
aggressive approach, 271
algorithmic trading, 134
alternative data, 131
alternative hypothesis, 86
amortizing bond, 22
analysis of variance (ANOVA), 119
annual general meeting, 257
annualized return, 10
annuity, 23
arithmetic mean, 5, 35
arithmetic mean return, 3
articles of incorporation, 240
artificial intelligence, 132
asymmetric information, 299
auction pricing, 304
audit committee, 259
autarky, 199
automatic stabilizers, 174

B

B2B, 304
B2C, 304
base currency, 218

Bayes' formula, 56
biased estimator, 39
Big Data, 131
bilateralism, 199
bimodal, 36
binomial model, 31
black swan risk, 201
board of directors, 240, 258
bondholder, 249
bond indenture, 258
bonding costs, 299
bond market vigilantes, 191
bootstrap, 81
bootstrap resampling, 73
borrowed capital, 240
box and whisker plot, 37
breakeven point, 140
budget deficit, 173
budget surplus, 173
bundling, 304
business cycle, 163
business model, 303
buy side, 218

C

capital allocation process, 276
capital restrictions, 211
capital spending, 176
capital structure irrelevance, 292
cartel, 151
cash conversion cycle (CCC), 265
cash flow additivity principle, 27
central limit theorem, 79
classical cycle, 163
channel strategy, 304
cluster sampling, 78
coefficient of determination, 121
coefficient of variation (CV), 40
coincident indicators, 168
collateral, 258

commodity producers, 304
common stock, 24
compensation committee, 259
compounding frequency, 11
concentration measures, 155
conditional expected value, 55
conservative approach, 270
constant growth DDM, 24
consumer spending, 165
contraction, 163
contractionary monetary policy, 173
controlling shareholder, 256
convenience sampling, 79
conventional cash flow pattern, 279
conventional fixed peg arrangement, 222
convertible debt, 291
cooperative, 197
corporate exhaust, 132
corporate governance, 257
corporation, 240
corporations, 218
correlation, 64
correlation coefficient, 48
cost averaging, 5
cost of liquidity, 268
costs of asymmetric information, 299
costs of financial distress, 295
coupon rate, 22
Cournot model, 150
covariance, 47, 61
covariance matrix, 62
covenants, 258
crawling peg, 223
credit cycles, 164
creditor committee, 258
critical value, 87
cross rate, 229
crowding-out effect, 175, 179
crowdsourcing, 306
currency board arrangement, 222
current spending, 176
customer, 250

cyclically adjusted budget deficit, 180

D

data science, 132
days inventory on hand, 265
days payable outstanding, 266
days sales outstanding, 266
debt ratio, 174
decile, 37
decision rule, 87, 89
deep learning, 133
default risk, 2
demographic factors, 166
dependent variable, 111
differences between means, 91
dilution, 249
direct listing, 242
direct quote, 218
direct sales, 304
direct taxes, 176
disadvantages of the IRR method, 279
discount rate, 1
discretionary fiscal policy, 174, 178
diseconomies of scale, 143
dispersion, 38
disposable income, 177
dividend discount model (DDM), 24
dividends, 241
dividend yield, 27
dominant firm model, 151
double taxation, 240
downside risk, 41
drag on liquidity, 269
dual-class structure, 256
durable goods, 166
dynamic pricing, 304

E

economic tools, 202
employee stock ownership plan (ESOP), 259
equality of variances, 99
equity securities, 23
event risk, 201
excess kurtosis, 45
exchange rate, 218
exchange rate regimes, 222
exchange rate targeting, 190
exogenous risk, 201
expansion, 163
expansionary monetary policy, 173
expansion options, 283
expansion projects, 275
expected value, 53
export subsidy, 209, 210
extraordinary general meeting, 258

F

F-distribution, 99
fiat money, 185
financial distress, 295
financial tools, 202
fintech, 131
fiscal multiplier, 177
fiscal policy, 173
fiscal policy tools, 175
fixed-coupon bond, 22
flexibility options, 284
foreign direct investment, 199
formal dollarization, 222
forward currency contract, 217
forward discount, 233
forward exchange rate, 220
forward premium, 233
franchising, 305
free cash flow hypothesis, 299
free float., 241
freemium pricing, 305
fundamental options, 284

G

general partnership, 239
geometric mean, 5
geometric mean return, 3
geophysical resource endowment, 198
geopolitical risk, 201
geopolitics, 197
globalization, 198
going concern projects, 275
Gordon growth model, 24
government, 251
government entities, 218
gross return, 12
growth cycle, 163
growth rate cycle, 164

H

harmonic mean, 4, 5
hedging, 217
hegemony, 199
Herfindahl-Hirschman Index (HHI), 155
heteroskedasticity, 117
hidden revenue, 305
high-frequency trading, 134
holding period return (HPR), 2
homoskedasticity, 117
hostile takeover, 258
housing costs relative to income, 166
human capital, 250
hurdle rate, 278
hypothesis, 85
hypothesis testing steps, 85

I

impact lag, 179
impact of geopolitical risk, 202
implied forward rate, 29

implied growth rate, 27
independent director, 250
independently and identically distributed., 72
independently floating exchange rate, 223
independent variable, 111
indirect quote, 218
indirect taxes, 176
inflation premium, 2
inflation reports, 190
inflation targeting, 190
information asymmetry, 256
initial public offering (IPO), 242
inside director, 250
institutions, 198
interest rate targeting, 190
internal rate of return (IRR), 6, 278, 279
  disadvantages, 279
  IRR decision rule, 278
International Monetary Fund (IMF), 200
Internet of Things, 132
interpretation of the skewness measure, 45
interquartile range, 37
inventory-sales ratio, 165
investment accounts, 218

J

jackknife, 81
judgmental sampling, 79

K

key advantage of IRR, 279
key advantage of NPV, 279
kinked demand curve model, 148
kurtosis, 45

L

labor, 165
lagging indicators, 168
latency, 132
leading indicators, 168
legal tender, 185
leptokurtic, 45
leveraged accounts, 218
leveraged return, 13
licensing, 305
licensing agreements, 306
likelihood of geopolitical risk, 201
limited company, 240
limited liability, 240
limited liability partnership (LLP), 240
limited partner, 240
limited partnership, 240
lin-log model, 126
liquidity, 267
liquidity risk, 2
liquidity sources, 267
liquidity trap, 191
listed companies, 241
location of the mean, median, and mode, 43
log-lin model, 126
log-log model, 126
lognormal distribution, 71
long run, 139
long-run shutdown point, 140

M

machine learning, 132
managed floating exchange rates, 223
management board, 250
management of exchange rates within crawling bands, 223
marginal propensity to consume (MPC), 177
match funding, 275
material environmental factors, 251
maturity risk, 2
mean absolute deviation (MAD), 38
mean differences, 94

mean squared error (MSE), 120
mean square regression (MSR), 119
measures of central tendency, 35
median, 35
menu costs, 186
mesokurtic, 45
minimum domestic content requirement, 209
minimum efficient scale, 143
minority shareholder, 256
MM I with taxes, 295
MM Proposition I (no taxes), 292
modal interval, 36
mode, 36
moderate approach, 271
monetary policy, 173
monetary policy tools, 186
monetary transmission mechanism, 187
monetary union, 222
money-weighted rate of return, 6
money-weighted return, 6
monopolistic competition, 145
monopoly, 145
monitoring costs, 299
Monte Carlo simulation, 72
mortgage rates, 166
multilateralism, 199
multistage DDM, 25

N

Nash equilibrium, 150
nationalism, 198
national security tools, 202
natural language processing, 133
negative externalities, 251
negative skew, 43
net agency cost of equity, 299
net operating profit after tax (NOPAT), 280
net present value (NPV), 6, 277
net return, 12
network effects, 306

net working capital, 267
neural networks, 132
neutral interest rate, 190
N-firm concentration ratio, 155
no-arbitrage principle, 28
nominal exchange rate, 219
nominating/governance committee, 259
noncooperative, 197
nondurable goods, 166
nonparametric test, 101
nonprobability sampling, 77
nonstate actors, 197
null hypothesis, 86

O

objective of a central bank, 186
oligopoly, 145
omnichannel strategy, 304
one-stage cluster sampling, 78
one-tier board structure, 250
operational independence, 189
optimal capital structure, 296
optional products, 304
ordinary least squares (OLS), 113
outlier, , 35, 43
overfitting, 133
ownership capital, 240

P

paired comparisons test, 93, 94
parametric test, 101
partnership agreement, 240
par value, 23
peak, 163
pecking order theory, 300
pegging, 186
penetration pricing, 304
percentile, 37

perfect competition, 145
perpetual bond, 22
perpetuity, 22
physical capital, 165
physical risk, 251
platykurtic, 45
portfolio investment flows, 199
portfolio variance, 63
positive skew, 43
power of a test, 89
predicted values, 124
preferred stock, 23
premium, 21
pretax nominal return, 12
price currency, 218
price discrimination, 304
price relative, 12
price-setting options, 284
primary liquidity, 267
principal-agent conflict, 255
private debtholder, 249
private label manufacturers, 305
private limited company, 240
private placements, 242
probability of financial distress, 296
probability sampling, 77
probability tree, 55
production-flexibility options, 284
properties of covariance, 62
proxy, 257
proxy fight, 258
public corporation, 240
public limited company, 240
pull on liquidity, 269
pure discount, 21
p-value, 90

Q

quantile, 37
quantitative easing, 192

quartile, 37
quintile, 37
quota, 209
quota rents, 210

R

random sampling, 77
range, 38
razors-and-blades, 304
real exchange rate, 219
real money accounts, 218
real options, 283
real return, 12
real risk-free rate, 1
recession, 163
recognition lag, 179
recovery, 164
regionalism, 199
regional trading agreement (RTA), 212
regression line, 113
regulatory/compliance projects, 275
relative dispersion, 40
remuneration committee, 259
replication, 28
required rate of return, 1
required return, 23
resampling, 73
residual, 112
residual losses, 299
retail FX market, 218
return on capital, 280
return on invested capital (ROIC), 280
revenue tools, 175
Ricardian equivalence, 178
roles of central banks, 185
Roy's safety-first criterion, 65

S

sample covariance, 47, 62
sample kurtosis, 46
sample mean, 35
sample skewness, 44
sample standard deviation, 40
sample variance, 39
sanctions, 202
scatter plots, 46
scenario analysis, 202
secondary liquidity sources, 267
segmenting, 303
sell side, 218
senior manager, 250
services, 166
shareholder, 249
shareholder theory, 249
shoe leather costs, 186
shortfall risk, 65
short run, 139
short-run shutdown point, 140
significance level, 88
signposts, 202
simple linear regression, 111
simple random sampling, 77, 78
skew, skewness, 43
slope coefficient, 113
soft power, 198
sole proprietorship, 239
sovereign wealth funds, 218
Spearman rank correlation test, 106
special purpose acquisition company (SPAC), 242
speculative activity, 166
spending tools, 175
spot exchange rate, 220
spurious correlation, 49
Stackelberg model, 150
staggered board, 250
stakeholder management, 257
stakeholder theory, 249
standard deviation, 40
standard error of the sample mean, 80
standardization, 198

state actors, 197
static tradeoff theory, 296
stranded asset, 251
stratified random sampling, 78
structural budget deficit, 180
subscription model, 305
sum of squared errors (SSE), 113, 120
sum of squares regression (SSR), 119, 120
sunk costs, 282
supervised learning, 133
supervisory board, 250
supplier, 250
symmetrical, 43
systematic sampling, 78

T

target capital structure, 298
target downside deviation, 41
target independence, 189
target semideviation, 41
target zone, 222
tariff, 209
tax shield, 294
tax shield provided by debt, 295
tender offer, 258
test statistic, 88
text analytics, 133
thematic risk, 201
threshold level, 65
tiered pricing, 304
time preference, 1
time-weighted rate of return, 8
timing options, 283
tools of geopolitics, 202
total sum of squares (SST), 119
total working capital, 267
trading blocs, 212
transfer payments, 175
trimmed mean, 37
trimmed or winsorized mean, 5

trimodal, 36
trough, 163
t-test, 123
two-stage cluster sampling, 78
two-tier board structure, 250
Type I error, 88
Type II error, 88

U

unconventional cash flow pattern, 279
underfitting, 133
unimodal, 36
unsupervised learning, 133

V

value-added resellers, 306
value-based pricing, 304
value chain, 305
value of a levered firm, 295
value of an unlevered firm, 295
value proposition, 305
variance (from a probability model), 54
velocity of geopolitical risk, 201
volatility, 54
voluntary export restraint, 209
voluntary export restraint (VER), 210
voting rights, 241

W

weighted-average cost of capital (WACC), 289
winsorized mean, 37
World Bank, 200
World Trade Organization (WTO), 200

Y

yield to maturity, 21

Z

zero-coupon bond, 21