CFA Institute

CFA INSTITUTE INVESTMENT SERIES

Fourth Edition

Quantitative Investment Analysis

WILEY

QUANTITATIVE INVESTMENT ANALYSIS

CFA Institute is the premier association for investment professionals around the world, with over 160,000 members in 165 countries and territories. Since 1963 the organization has developed and administered the renowned Chartered Financial Analyst® Program. With a rich history of leading the investment profession, CFA Institute has set the highest standards in ethics, education, and professional excellence within the global investment community and is the foremost authority on investment profession conduct and practice. Each book in the CFA Institute Investment Series is geared toward industry practitioners along with graduate-level finance students and covers the most important topics in the industry. The authors of these cutting-edge books are themselves industry professionals and academics and bring their wealth of knowledge and expertise to this series.

QUANTITATIVE INVESTMENT ANALYSIS

Fourth Edition

Richard A. DeFusco, CFA

Dennis W. McLeavey, CFA

Jerald E. Pinto, CFA

David E. Runkle, CFA

WILEY

Cover image: © r.nagy/Shutterstock
Cover design: Wiley

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

ISBN 978-1-119-74362-0 (Hardcover)
ISBN 978-1-119-74365-1 (ePDF)
ISBN 978-1-119-74364-4 (ePub)

Printed in the United States of America.
10 9 8 7 6 5 4 3 2 1

CONTENTS

CHAPTER 2
Organizing, Visualizing, and Describing Data 45

CHAPTER 8
Multiple Regression 365

CHAPTER 9
Time-Series Analysis 451

PREFACE

We are pleased to bring you *Quantitative Investment Analysis, Fourth Edition*, which focuses on key tools that are needed for today's professional investor. In addition to classic areas such as the time value of money and probability and statistics, the text covers advanced concepts in regression, time series, machine learning, and big data projects. The text teaches critical skills that challenge many professionals, and shows how these techniques can be applied to areas such as factor modeling, risk management, and backtesting and simulation.

The content was developed in partnership by a team of distinguished academics and practitioners, chosen for their acknowledged expertise in the field, and guided by CFA Institute. It is written specifically with the investment practitioner in mind and is replete with examples and practice problems that reinforce the learning outcomes and demonstrate real-world applicability.

The CFA Program Curriculum, from which the content of this book was drawn, is subjected to a rigorous review process to assure that it is:

- Faithful to the findings of our ongoing industry practice analysis
- Valuable to members, employers, and investors
- Globally relevant
- Generalist (as opposed to specialist) in nature
- Replete with sufficient examples and practice opportunities
- Pedagogically sound

The accompanying workbook is a useful reference that provides Learning Outcome Statements that describe exactly what readers will learn and be able to demonstrate after mastering the accompanying material. Additionally, the workbook has summary overviews and practice problems for each chapter.

We are confident that you will find this and other books in the CFA Institute Investment Series helpful in your efforts to grow your investment knowledge, whether you are a relatively new entrant or an experienced veteran striving to keep up to date in the ever-changing market environment. CFA Institute, as a long-term committed participant in the investment profession and a not-for-profit global membership association, is pleased to provide you with this opportunity.

ACKNOWLEDGMENTS

Special thanks to all the reviewers, advisors, and question writers who helped to ensure high practical relevance, technical correctness, and understandability of the material presented here.

We would like to thank the many others who played a role in the conception and production of this book: the Curriculum and Learning Experience team at CFA Institute, with special thanks to the curriculum directors, past and present, who worked with the authors and reviewers to produce the chapters in this book; the Practice Analysis team at CFA Institute; and the Publishing and Technology team for bringing this book to production.

ABOUT THE CFA INSTITUTE INVESTMENT SERIES

CFA Institute is pleased to provide the CFA Institute Investment Series, which covers major areas in the field of investments. We provide this best-in-class series for the same reason we have been chartering investment professionals for more than 45 years: to lead the investment profession globally by setting the highest standards of ethics, education, and professional excellence.

The books in the CFA Institute Investment Series contain practical, globally relevant material. They are intended both for those contemplating entry into the extremely competitive field of investment management as well as for those seeking a means of keeping their knowledge fresh and up to date. This series was designed to be user friendly and highly relevant.

We hope you find this series helpful in your efforts to grow your investment knowledge, whether you are a relatively new entrant or an experienced veteran ethically bound to keep up to date in the ever-changing market environment. As a long-term, committed participant in the investment profession and a not-for-profit global membership association, CFA Institute is pleased to provide you with this opportunity.

THE TEXTS

Corporate Finance: A Practical Approach is a solid foundation for those looking to achieve lasting business growth. In today's competitive business environment, companies must find innovative ways to enable rapid and sustainable growth. This text equips readers with the foundational knowledge and tools for making smart business decisions and formulating strategies to maximize company value. It covers everything from managing relationships between stakeholders to evaluating merger and acquisition bids, as well as the companies behind them. Through extensive use of real-world examples, readers will gain critical perspective into interpreting corporate financial data, evaluating projects, and allocating funds in ways that increase corporate value. Readers will gain insights into the tools and strategies used in modern corporate financial management.

Equity Asset Valuation is a particularly cogent and important resource for anyone involved in estimating the value of securities and understanding security pricing. A well-informed professional knows that the common forms of equity valuation—dividend discount modeling, free cash flow modeling, price/earnings modeling, and residual income modeling—can all be reconciled with one another under certain assumptions. With a deep understanding of the underlying assumptions, the professional investor can better understand what other investors assume when calculating their valuation estimates. This text has a global orientation, including emerging markets.

Fixed Income Analysis has been at the forefront of new concepts in recent years, and this particular text offers some of the most recent material for the seasoned professional who is not a fixed-income specialist. The application of option and derivative technology to the once staid province of fixed income has helped contribute to an explosion of thought in this area. Professionals have been challenged to stay up to speed with credit derivatives, swaptions, collateralized mortgage securities, mortgage-backed securities, and other vehicles, and this explosion of products has strained the world's financial markets and tested central banks to provide sufficient oversight. Armed with a thorough grasp of the new exposures, the professional investor is much better able to anticipate and understand the challenges our central bankers and markets face.

International Financial Statement Analysis is designed to address the ever-increasing need for investment professionals and students to think about financial statement analysis from a global perspective. The text is a practically oriented introduction to financial statement analysis that is distinguished by its combination of a true international orientation, a structured presentation style, and abundant illustrations and tools covering concepts as they are introduced in the text. The authors cover this discipline comprehensively and with an eye to ensuring the reader's success at all levels in the complex world of financial statement analysis.

Investments: Principles of Portfolio and Equity Analysis provides an accessible yet rigorous introduction to portfolio and equity analysis. Portfolio planning and portfolio management are presented within a context of up-to-date, global coverage of security markets, trading, and market-related concepts and products. The essentials of equity analysis and valuation are explained in detail and profusely illustrated. The book includes coverage of practitioner-important but often neglected topics, such as industry analysis. Throughout, the focus is on the practical application of key concepts with examples drawn from both emerging and developed markets. Each chapter affords the reader many opportunities to self-check his or her understanding of topics.

All books in the CFA Institute Investment Series are available through all major booksellers. And, all titles are available on the Wiley Custom Select platform at http://customselect.wiley.com/ where individual chapters for all the books may be mixed and matched to create custom textbooks for the classroom.

CHAPTER 1

THE TIME VALUE OF MONEY

Richard A. DeFusco, PhD, CFA
Dennis W. McLeavey, DBA, CFA
Jerald E. Pinto, PhD, CFA
David E. Runkle, PhD, CFA

LEARNING OUTCOMES

The candidate should be able to:

- interpret interest rates as required rates of return, discount rates, or opportunity costs;
- explain an interest rate as the sum of a real risk-free rate and premiums that compensate investors for bearing distinct types of risk;
- calculate and interpret the effective annual rate, given the stated annual interest rate and the frequency of compounding;
- solve time value of money problems for different frequencies of compounding;
- calculate and interpret the future value (FV) and present value (PV) of a single sum of money, an ordinary annuity, an annuity due, a perpetuity (PV only), and a series of unequal cash flows;
- demonstrate the use of a time line in modeling and solving time value of money problems.

1. INTRODUCTION

As individuals, we often face decisions that involve saving money for a future use, or borrowing money for current consumption. We then need to determine the amount we need to invest, if we are saving, or the cost of borrowing, if we are shopping for a loan. As investment analysts, much of our work also involves evaluating transactions with present and future cash flows. When we place a value on any security, for example, we are attempting to determine the worth of a stream of future cash flows. To carry out all the above tasks accurately, we must understand the mathematics of time value of money problems. Money has time value in that individuals value a given amount of money more highly the earlier it is received. Therefore, a smaller amount of money now may be equivalent in value to a larger amount received at a future date. The **time value of money** as a topic in investment mathematics deals with equivalence relationships between cash flows with different dates. Mastery of time value of money concepts and techniques is essential for investment analysts.

The chapter[1] is organized as follows: Section 2 introduces some terminology used throughout the chapter and supplies some economic intuition for the variables we will discuss. Section 3 tackles the problem of determining the worth at a future point in time of an amount invested today. Section 4 addresses the future worth of a series of cash flows. These two sections provide the tools for calculating the equivalent value at a future date of a single cash flow or series of cash flows. Sections 5 and 6 discuss the equivalent value today of a single future cash flow and a series of future cash flows, respectively. In Section 7, we explore how to determine other quantities of interest in time value of money problems.

2. INTEREST RATES: INTERPRETATION

In this chapter, we will continually refer to interest rates. In some cases, we assume a particular value for the interest rate; in other cases, the interest rate will be the unknown quantity we seek to determine. Before turning to the mechanics of time value of money problems, we must illustrate the underlying economic concepts. In this section, we briefly explain the meaning and interpretation of interest rates.

Time value of money concerns equivalence relationships between cash flows occurring on different dates. The idea of equivalence relationships is relatively simple. Consider the following exchange: You pay \$10,000 today and in return receive \$9,500 today. Would you accept this arrangement? Not likely. But what if you received the \$9,500 today and paid the \$10,000 one year from now? Can these amounts be considered equivalent? Possibly, because a payment of \$10,000 a year from now would probably be worth less to you than a payment of \$10,000 today. It would be fair, therefore, to **discount** the \$10,000 received in one year; that is, to cut its value based on how much time passes before the money is paid. An **interest rate**, denoted *r*, is a rate of return that reflects the relationship between differently dated cash flows. If \$9,500 today and \$10,000 in one year are equivalent in value, then \$10,000 − \$9,500 = \$500 is the required compensation for receiving \$10,000 in one year rather than now. The interest rate—the required compensation stated as a rate of return—is \$500/\$9,500 = 0.0526 or 5.26 percent.

Interest rates can be thought of in three ways. First, they can be considered required rates of return—that is, the minimum rate of return an investor must receive in order to accept the investment. Second, interest rates can be considered discount rates. In the example above, 5.26 percent is that rate at which we discounted the \$10,000 future amount to find its value today. Thus, we use the terms "interest rate" and "discount rate" almost interchangeably. Third, interest rates can be considered opportunity costs. An **opportunity cost** is the value that investors forgo by choosing a particular course of action. In the example, if the party who supplied \$9,500 had instead decided to spend it today, he would have forgone earning

[1]Examples in this, and other chapters, in the text were updated by Professor Sanjiv Sabherwal of the University of Texas, Arlington.

5.26 percent on the money. So we can view 5.26 percent as the opportunity cost of current consumption.

Economics tells us that interest rates are set in the marketplace by the forces of supply and demand, where investors are suppliers of funds and borrowers are demanders of funds. Taking the perspective of investors in analyzing market-determined interest rates, we can view an interest rate r as being composed of a real risk-free interest rate plus a set of four premiums that are required returns or compensation for bearing distinct types of risk:

$$r = \text{Real risk-free interest rate} + \text{Inflation premium} + \text{Default risk premium} + \text{Liquidity premium} + \text{Maturity premium}$$

- The **real risk-free interest rate** is the single-period interest rate for a completely risk-free security if no inflation were expected. In economic theory, the real risk-free rate reflects the time preferences of individuals for current versus future real consumption.
- The **inflation premium** compensates investors for expected inflation and reflects the average inflation rate expected over the maturity of the debt. Inflation reduces the purchasing power of a unit of currency—the amount of goods and services one can buy with it. The sum of the real risk-free interest rate and the inflation premium is the **nominal risk-free interest rate**.[2] Many countries have governmental short-term debt whose interest rate can be considered to represent the nominal risk-free interest rate in that country. The interest rate on a 90-day US Treasury bill (T-bill), for example, represents the nominal risk-free interest rate over that time horizon.[3] US T-bills can be bought and sold in large quantities with minimal transaction costs and are backed by the full faith and credit of the US government.
- The **default risk premium** compensates investors for the possibility that the borrower will fail to make a promised payment at the contracted time and in the contracted amount.
- The **liquidity premium** compensates investors for the risk of loss relative to an investment's fair value if the investment needs to be converted to cash quickly. US T-bills, for example, do not bear a liquidity premium because large amounts can be bought and sold without affecting their market price. Many bonds of small issuers, by contrast, trade infrequently after they are issued; the interest rate on such bonds includes a liquidity premium reflecting the relatively high costs (including the impact on price) of selling a position.
- The **maturity premium** compensates investors for the increased sensitivity of the market value of debt to a change in market interest rates as maturity is extended, in general (holding all else equal). The difference between the interest rate on longer-maturity, liquid Treasury debt and that on short-term Treasury debt reflects a positive maturity premium for the longer-term debt (and possibly different inflation premiums as well).

---

[2]Technically, 1 plus the nominal rate equals the product of 1 plus the real rate and 1 plus the inflation rate. As a quick approximation, however, the nominal rate is equal to the real rate plus an inflation premium. In this discussion we focus on approximate additive relationships to highlight the underlying concepts.

[3]Other developed countries issue securities similar to US Treasury bills. The French government issues BTFs or negotiable fixed-rate discount Treasury bills (*Bons du Trésor àtaux fixe et à intérêts précomptés*) with maturities of up to one year. The Japanese government issues a short-term Treasury bill with maturities of 6 and 12 months. The German government issues at discount both Treasury financing paper (*Finanzierungsschätze des Bundes* or, for short, *Schätze*) and Treasury discount paper (*Bubills*) with maturities up to 24 months. In the United Kingdom, the British government issues gilt-edged Treasury bills with maturities ranging from 1 to 364 days. The Canadian government bond market is closely related to the US market; Canadian Treasury bills have maturities of 3, 6, and 12 months.

Using this insight into the economic meaning of interest rates, we now turn to a discussion of solving time value of money problems, starting with the future value of a single cash flow.

3. THE FUTURE VALUE OF A SINGLE CASH FLOW

In this section, we introduce time value associated with a single cash flow or lump-sum investment. We describe the relationship between an initial investment or **present value (PV)**, which earns a rate of return (the interest rate per period) denoted as r, and its **future value (FV)**, which will be received N years or periods from today.

The following example illustrates this concept. Suppose you invest \$100 (PV = \$100) in an interest-bearing bank account paying 5 percent annually. At the end of the first year, you will have the \$100 plus the interest earned, $0.05 \times \$100 = \$5$, for a total of \$105. To formalize this one-period example, we define the following terms:

$$\begin{aligned} \text{PV} &= \text{present value of the investment} \\ \text{FV}_N &= \text{future value of the investment } N \text{ periods from today} \\ r &= \text{rate of interest per period} \end{aligned}$$

For $N = 1$, the expression for the future value of amount PV is

$$\text{FV}_1 = \text{PV}(1 + r) \tag{1}$$

For this example, we calculate the future value one year from today as $\text{FV}_1 = \$100(1.05) = \$105$.

Now suppose you decide to invest the initial \$100 for two years with interest earned and credited to your account annually (annual compounding). At the end of the first year (the beginning of the second year), your account will have \$105, which you will leave in the bank for another year. Thus, with a beginning amount of \$105 (PV = \$105), the amount at the end of the second year will be \$105(1.05) = \$110.25. Note that the \$5.25 interest earned during the second year is 5 percent of the amount invested at the beginning of Year 2.

Another way to understand this example is to note that the amount invested at the beginning of Year 2 is composed of the original \$100 that you invested plus the \$5 interest earned during the first year. During the second year, the original principal again earns interest, as does the interest that was earned during Year 1. You can see how the original investment grows:

Original investment	\$100.00
Interest for the first year (\$100 × 0.05)	5.00
Interest for the second year based on original investment (\$100 × 0.05)	5.00
Interest for the second year based on interest earned in the first year (0.05 × \$5.00 interest on interest)	0.25
Total	\$110.25

The \$5 interest that you earned each period on the \$100 original investment is known as **simple interest** (the interest rate times the principal). **Principal** is the amount of funds originally invested. During the two-year period, you earn \$10 of simple interest. The extra \$0.25 that you have at the end of Year 2 is the interest you earned on the Year 1 interest of \$5 that you reinvested.

The interest earned on interest provides the first glimpse of the phenomenon known as **compounding**. Although the interest earned on the initial investment is important, for a given interest rate it is fixed in size from period to period. The compounded interest earned on reinvested interest is a far more powerful force because, for a given interest rate, it grows in size each period. The importance of compounding increases with the magnitude of the interest rate. For example, \$100 invested today would be worth about \$13,150 after 100 years if compounded annually at 5 percent, but worth more than \$20 million if compounded annually over the same time period at a rate of 13 percent.

To verify the \$20 million figure, we need a general formula to handle compounding for any number of periods. The following general formula relates the present value of an initial investment to its future value after N periods:

$$\mathrm{FV}_N = \mathrm{PV}(1 + r)^N \tag{2}$$

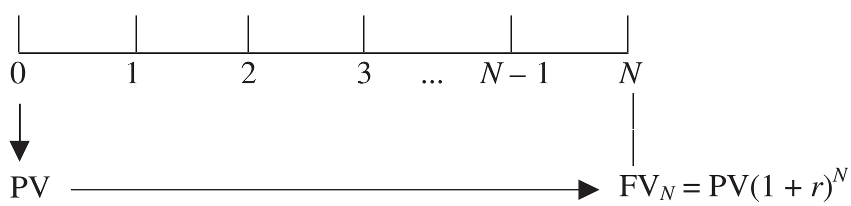
where r is the stated interest rate per period and N is the number of compounding periods. In the bank example, $\mathrm{FV}_2 = \$100(1 + 0.05)^2 = \$110.25$. In the 13 percent investment example, $\mathrm{FV}_{100} = \$100(1.13)^{100} = \$20{,}316{,}287.42$.

The most important point to remember about using the future value equation is that the stated interest rate, r, and the number of compounding periods, N, must be compatible. Both variables must be defined in the same time units. For example, if N is stated in months, then r should be the one-month interest rate, unannualized.

A time line helps us to keep track of the compatibility of time units and the interest rate per time period. In the time line, we use the time index t to represent a point in time a stated number of periods from today. Thus the present value is the amount available for investment today, indexed as $t = 0$. We can now refer to a time N periods from today as $t = N$. The time line in Figure 1 shows this relationship. In Figure 1, we have positioned the initial investment, PV, at $t = 0$. Using Equation 2, we move the present value, PV, forward to $t = N$ by the factor $(1 + r)^N$. This factor is called a future value factor. We denote the future value on the time line as FV and position it at $t = N$. Suppose the future value is to be received exactly 10 periods from today's date ($N = 10$). The present value, PV, and the future value, FV, are separated in time through the factor $(1 + r)^{10}$.

The fact that the present value and the future value are separated in time has important consequences:

FIGURE 1 The Relationship between an Initial Investment, PV, and Its Future Value, FV



- We can add amounts of money only if they are indexed at the same point in time.
- For a given interest rate, the future value increases with the number of periods.
- For a given number of periods, the future value increases with the interest rate.

To better understand these concepts, consider three examples that illustrate how to apply the future value formula.

EXAMPLE 1 The Future Value of a Lump Sum with Interim Cash Reinvested at the Same Rate

You are the lucky winner of your state's lottery of $5 million after taxes. You invest your winnings in a five-year certificate of deposit (CD) at a local financial institution. The CD promises to pay 7 percent per year compounded annually. This institution also lets you reinvest the interest at that rate for the duration of the CD. How much will you have at the end of five years if your money remains invested at 7 percent for five years with no withdrawals?

Solution: To solve this problem, compute the future value of the $5 million investment using the following values in Equation 2:

$$\begin{aligned} \text{PV} &= \$5{,}000{,}000 \\ r &= 7\% = 0.07 \\ N &= 5 \\ \text{FV}_N &= \text{PV}(1+r)^N \\ &= \$5{,}000{,}000(1.07)^5 \\ &= \$5{,}000{,}000(1.402552) \\ &= \$7{,}012{,}758.65 \end{aligned}$$

At the end of five years, you will have $7,012,758.65 if your money remains invested at 7 percent with no withdrawals.

Note that in this and most examples in this chapter, the factors are reported at six decimal places but the calculations may actually reflect greater precision. For example, the reported 1.402552 has been rounded up from 1.40255173 (the calculation is actually carried out with more than eight decimal places of precision by the calculator or spreadsheet). Our final result reflects the higher number of decimal places carried by the calculator or spreadsheet.[4]

[4]We could also solve time value of money problems using tables of interest rate factors. Solutions using tabled values of interest rate factors are generally less accurate than solutions obtained using calculators or spreadsheets, so practitioners prefer calculators or spreadsheets.

EXAMPLE 2 The Future Value of a Lump Sum with No Interim Cash

An institution offers you the following terms for a contract: For an investment of ¥2,500,000, the institution promises to pay you a lump sum six years from now at an 8 percent annual interest rate. What future amount can you expect?

Solution: Use the following data in Equation 2 to find the future value:

$$
\begin{aligned}
\text{PV} &= ¥2{,}500{,}000 \\
r &= 8\% = 0.08 \\
N &= 6 \\
\text{FV}_N &= \text{PV}(1+r)^N \\
&= ¥2{,}500{,}000(1.08)^6 \\
&= ¥2{,}500{,}000(1.586874) \\
&= ¥3{,}967{,}186
\end{aligned}
$$

You can expect to receive ¥3,967,186 six years from now.

Our third example is a more complicated future value problem that illustrates the importance of keeping track of actual calendar time.

EXAMPLE 3 The Future Value of a Lump Sum

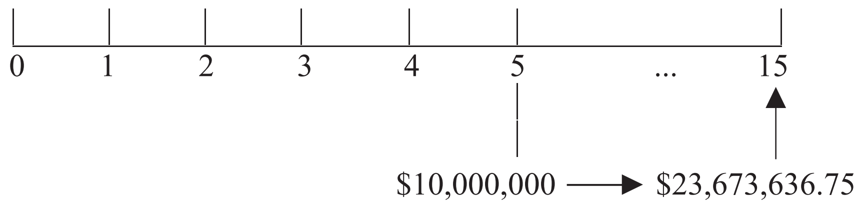
A pension fund manager estimates that his corporate sponsor will make a \$10 million contribution five years from now. The rate of return on plan assets has been estimated at 9 percent per year. The pension fund manager wants to calculate the future value of this contribution 15 years from now, which is the date at which the funds will be distributed to retirees. What is that future value?

Solution: By positioning the initial investment, PV, at $t = 5$, we can calculate the future value of the contribution using the following data in Equation 2:

$$
\begin{aligned}
\text{PV} &= \$10 \text{ million} \\
r &= 9\% = 0.09 \\
N &= 10 \\
\text{FV}_N &= \text{PV}(1+r)^N \\
&= \$10{,}000{,}000(1.09)^{10} \\
&= \$10{,}000{,}000(2.367364) \\
&= \$23{,}673{,}636.75
\end{aligned}
$$

This problem looks much like the previous two, but it differs in one important respect: its timing. From the standpoint of today ($t = 0$), the future amount of \$23,673,636.75 is 15 years into the future. Although the future value is 10 years from

FIGURE 2 The Future Value of a Lump Sum, Initial Investment Not at $t = 0$



its present value, the present value of $10 million will not be received for another five years.

As Figure 2 shows, we have followed the convention of indexing today as $t = 0$ and indexing subsequent times by adding 1 for each period. The additional contribution of $10 million is to be received in five years, so it is indexed as $t = 5$ and appears as such in the figure. The future value of the investment in 10 years is then indexed at $t = 15$; that is, 10 years following the receipt of the $10 million contribution at $t = 5$. Time lines like this one can be extremely useful when dealing with more-complicated problems, especially those involving more than one cash flow.

In a later section of this chapter, we will discuss how to calculate the value today of the $10 million to be received five years from now. For the moment, we can use Equation 2. Suppose the pension fund manager in **Example 3** above were to receive $6,499,313.86 today from the corporate sponsor. How much will that sum be worth at the end of five years? How much will it be worth at the end of 15 years?

$$\begin{aligned}
\text{PV} &= \$6{,}499{,}313.86 \\
r &= 9\% = 0.09 \\
N &= 5 \\
\text{FV}_N &= \text{PV}(1 + r)^N \\
&= \$6{,}499{,}313.86(1.09)^5 \\
&= \$6{,}499{,}313.86(1.538624) \\
&= \$10{,}000{,}000 \text{ at the five-year mark}
\end{aligned}$$

and

$$\begin{aligned}
\text{PV} &= \$6{,}499{,}313.86 \\
r &= 9\% = 0.09 \\
N &= 15
\end{aligned}$$

$$\begin{aligned}\mathrm{FV}_N &= \mathrm{PV}(1+r)^N \\ &= \$6{,}499{,}313.86(1.09)^{15} \\ &= \$6{,}499{,}313.86(3.642482) \\ &= \$23{,}673{,}636.74 \text{ at the 15-year mark}\end{aligned}$$

These results show that today's present value of about $6.5 million becomes $10 million after five years and $23.67 million after 15 years.

3.1. The Frequency of Compounding

In this section, we examine investments paying interest more than once a year. For instance, many banks offer a monthly interest rate that compounds 12 times a year. In such an arrangement, they pay interest on interest every month. Rather than quote the periodic monthly interest rate, financial institutions often quote an annual interest rate that we refer to as the **stated annual interest rate** or **quoted interest rate**. We denote the stated annual interest rate by r_s. For instance, your bank might state that a particular CD pays 8 percent compounded monthly. The stated annual interest rate equals the monthly interest rate multiplied by 12. In this example, the monthly interest rate is $0.08/12 = 0.0067$ or 0.67 percent.[5] This rate is strictly a quoting convention because $(1 + 0.0067)^{12} = 1.083$, not 1.08; the term $(1 + r_s)$ is not meant to be a future value factor when compounding is more frequent than annual.

With more than one compounding period per year, the future value formula can be expressed as

$$\mathrm{FV}_N = \mathrm{PV}\left(1 + \frac{r_s}{m}\right)^{mN} \tag{3}$$

where r_s is the stated annual interest rate, m is the number of compounding periods per year, and N now stands for the number of years. Note the compatibility here between the interest rate used, r_s/m, and the number of compounding periods, mN. The periodic rate, r_s/m, is the stated annual interest rate divided by the number of compounding periods per year. The number of compounding periods, mN, is the number of compounding periods in one year multiplied by the number of years. The periodic rate, r_s/m, and the number of compounding periods, mN, must be compatible.

EXAMPLE 4 The Future Value of a Lump Sum with Quarterly Compounding

Continuing with the CD example, suppose your bank offers you a CD with a two-year maturity, a stated annual interest rate of 8 percent compounded quarterly, and a feature

[5]To avoid rounding errors when using a financial calculator, divide 8 by 12 and then press the %*i* key, rather than simply entering 0.67 for %*i*, so we have $(1 + 0.08/12)^{12} = 1.083000$.

allowing reinvestment of the interest at the same interest rate. You decide to invest \$10,000. What will the CD be worth at maturity?

Solution: Compute the future value with Equation 3 as follows:

$$
\begin{aligned}
\text{PV} &= \$10{,}000 \\
r_s &= 8\% = 0.08 \\
m &= 4 \\
r_s/m &= 0.08/4 = 0.02 \\
N &= 2 \\
mN = 4(2) &= 8 \text{ interest periods} \\
\text{FV}_N &= \text{PV}\left(1 + \frac{r_s}{m}\right)^{mN} \\
&= \$10{,}000(1.02)^8 \\
&= \$10{,}000(1.171659) \\
&= \$11{,}716.59
\end{aligned}
$$

At maturity, the CD will be worth \$11,716.59.

The future value formula in Equation 3 does not differ from the one in Equation 2. Simply keep in mind that the interest rate to use is the rate per period and the exponent is the number of interest, or compounding, periods.

EXAMPLE 5 The Future Value of a Lump Sum with Monthly Compounding

An Australian bank offers to pay you 6 percent compounded monthly. You decide to invest A\$1 million for one year. What is the future value of your investment if interest payments are reinvested at 6 percent?

Solution: Use Equation 3 to find the future value of the one-year investment as follows:

$$
\begin{aligned}
\text{PV} &= \text{A}\$1{,}000{,}000 \\
r_s &= 6\% = 0.06 \\
m &= 12 \\
r_s/m &= 0.06/12 = 0.0050 \\
N &= 1 \\
mN &= 12(1) = 12 \text{ interest periods} \\
\text{FV}_N &= \text{PV}\left(1 + \frac{r_s}{m}\right)^{mN} \\
&= \text{A}\$1{,}000{,}000(1.005)^{12}
\end{aligned}
$$

$$= \text{A\$}1{,}000{,}000(1.061678)$$
$$= \text{A\$}1{,}061{,}677.81$$

If you had been paid 6 percent with annual compounding, the future amount would be only A$1,000,000(1.06) = A$1,060,000 instead of A$1,061,677.81 with monthly compounding.

3.2. Continuous Compounding

The preceding discussion on compounding periods illustrates discrete compounding, which credits interest after a discrete amount of time has elapsed. If the number of compounding periods per year becomes infinite, then interest is said to compound continuously. If we want to use the future value formula with continuous compounding, we need to find the limiting value of the future value factor for $m \to \infty$ (infinitely many compounding periods per year) in Equation 3. The expression for the future value of a sum in N years with continuous compounding is

$$\text{FV}_N = \text{PV}e^{r_s N} \tag{4}$$

The term $e^{r_s N}$ is the transcendental number $e \approx 2.7182818$ raised to the power $r_s N$. Most financial calculators have the function e^x.

EXAMPLE 6 The Future Value of a Lump Sum with Continuous Compounding

Suppose a $10,000 investment will earn 8 percent compounded continuously for two years. We can compute the future value with Equation 4 as follows:

$$\begin{aligned}
\text{PV} &= \$10{,}000 \\
r_s &= 8\% = 0.08 \\
N &= 2 \\
\text{FV}_N &= \text{PV}e^{r_s N} \\
&= \$10{,}000e^{0.08(2)} \\
&= \$10{,}000(1.173511) \\
&= \$11{,}735.11
\end{aligned}$$

With the same interest rate but using continuous compounding, the $10,000 investment will grow to $11,735.11 in two years, compared with $11,716.59 using quarterly compounding as shown in **Example 4**.

TABLE 1 The Effect of Compounding Frequency on Future Value

Frequency	r_s/m	mN	Future Value of \$1
Annual	$8\%/1 = 8\%$	$1 \times 1 = 1$	$\$1.00(1.08) = \$1.08$
Semiannual	$8\%/2 = 4\%$	$2 \times 1 = 2$	$\$1.00(1.04)^2 = \$1.081600$
Quarterly	$8\%/4 = 2\%$	$4 \times 1 = 4$	$\$1.00(1.02)^4 = \$1.082432$
Monthly	$8\%/12 = 0.6667\%$	$12 \times 1 = 12$	$\$1.00(1.006667)^{12} = \$1.083000$
Daily	$8\%/365 = 0.0219\%$	$365 \times 1 = 365$	$\$1.00(1.000219)^{365} = \$1.083278$
Continuous			$\$1.00e^{0.08(1)} = \$1.083287$

Table 1 shows how a stated annual interest rate of 8 percent generates different ending dollar amounts with annual, semiannual, quarterly, monthly, daily, and continuous compounding for an initial investment of \$1 (carried out to six decimal places).

As Table 1 shows, all six cases have the same stated annual interest rate of 8 percent; they have different ending dollar amounts, however, because of differences in the frequency of compounding. With annual compounding, the ending amount is \$1.08. More frequent compounding results in larger ending amounts. The ending dollar amount with continuous compounding is the maximum amount that can be earned with a stated annual rate of 8 percent.

Table 1 also shows that a \$1 investment earning 8.16 percent compounded annually grows to the same future value at the end of one year as a \$1 investment earning 8 percent compounded semiannually. This result leads us to a distinction between the stated annual interest rate and the **effective annual rate** (EAR).[6] For an 8 percent stated annual interest rate with semiannual compounding, the EAR is 8.16 percent.

3.3. Stated and Effective Rates

The stated annual interest rate does not give a future value directly, so we need a formula for the EAR. With an annual interest rate of 8 percent compounded semiannually, we receive a periodic rate of 4 percent. During the course of a year, an investment of \$1 would grow to $\$1(1.04)^2 = \$1.0816$, as illustrated in Table 1. The interest earned on the \$1 investment is \$0.0816 and represents an effective annual rate of interest of 8.16 percent. The effective annual rate is calculated as follows:

$$\text{EAR} = (1 + \text{Periodic interest rate})^m - 1 \tag{5}$$

[6]Among the terms used for the effective annual return on interest-bearing bank deposits are annual percentage yield (APY) in the United States and equivalent annual rate (EAR) in the United Kingdom. By contrast, the **annual percentage rate** (APR) measures the cost of borrowing expressed as a yearly rate. In the United States, the APR is calculated as a periodic rate times the number of payment periods per year and, as a result, some writers use APR as a general synonym for the stated annual interest rate. Nevertheless, APR is a term with legal connotations; its calculation follows regulatory standards that vary internationally. Therefore, “stated annual interest rate” is the preferred general term for an annual interest rate that does not account for compounding within the year.

The periodic interest rate is the stated annual interest rate divided by m, where m is the number of compounding periods in one year. Using our previous example, we can solve for EAR as follows: $(1.04)^2 - 1 = 8.16$ percent.

The concept of EAR extends to continuous compounding. Suppose we have a rate of 8 percent compounded continuously. We can find the EAR in the same way as above by finding the appropriate future value factor. In this case, a \$1 investment would grow to $\$1e^{0.08(1.0)} = \$1.0833$. The interest earned for one year represents an effective annual rate of 8.33 percent and is larger than the 8.16 percent EAR with semiannual compounding because interest is compounded more frequently. With continuous compounding, we can solve for the effective annual rate as follows:

$$\text{EAR} = e^{r_s} - 1 \tag{6}$$

We can reverse the formulas for EAR with discrete and continuous compounding to find a periodic rate that corresponds to a particular effective annual rate. Suppose we want to find the appropriate periodic rate for a given effective annual rate of 8.16 percent with semiannual compounding. We can use Equation 5 to find the periodic rate:

$$\begin{aligned}
0.0816 &= (1 + \text{Periodic rate})^2 - 1 \\
1.0816 &= (1 + \text{Periodic rate})^2 \\
(1.0816)^{1/2} - 1 &= \text{Periodic rate} \\
(1.04) - 1 &= \text{Periodic rate} \\
4\% &= \text{Periodic rate}
\end{aligned}$$

To calculate the continuously compounded rate (the stated annual interest rate with continuous compounding) corresponding to an effective annual rate of 8.33 percent, we find the interest rate that satisfies Equation 6:

$$\begin{aligned}
0.0833 &= e^{r_s} - 1 \\
1.0833 &= e^{r_s}
\end{aligned}$$

To solve this equation, we take the natural logarithm of both sides. (Recall that the natural log of e^{r_s} is $\ln e^{r_s} = r_s$.) Therefore, $\ln 1.0833 = r_s$, resulting in $r_s = 8$ percent. We see that a stated annual rate of 8 percent with continuous compounding is equivalent to an EAR of 8.33 percent.

4. THE FUTURE VALUE OF A SERIES OF CASH FLOWS

In this section, we consider series of cash flows, both even and uneven. We begin with a list of terms commonly used when valuing cash flows that are distributed over many time periods.

- An **annuity** is a finite set of level sequential cash flows.
- An **ordinary annuity** has a first cash flow that occurs one period from now (indexed at $t = 1$).

- An **annuity due** has a first cash flow that occurs immediately (indexed at $t = 0$).
- A **perpetuity** is a perpetual annuity, or a set of level never-ending sequential cash flows, with the first cash flow occurring one period from now.

4.1. Equal Cash Flows—Ordinary Annuity

Consider an ordinary annuity paying 5 percent annually. Suppose we have five separate deposits of $1,000 occurring at equally spaced intervals of one year, with the first payment occurring at $t = 1$. Our goal is to find the future value of this ordinary annuity after the last deposit at $t = 5$. The increment in the time counter is one year, so the last payment occurs five years from now. As the time line in Figure 3 shows, we find the future value of each $1,000 deposit as of $t = 5$ with Equation 2, $\text{FV}_N = \text{PV}(1 + r)^N$. The arrows in Figure 3 extend from the payment date to $t = 5$. For instance, the first $1,000 deposit made at $t = 1$ will compound over four periods. Using Equation 2, we find that the future value of the first deposit at $t = 5$ is $\$1{,}000(1.05)^4 = \$1{,}215.51$. We calculate the future value of all other payments in a similar fashion. (Note that we are finding the future value at $t = 5$, so the last payment does not earn any interest.) With all values now at $t = 5$, we can add the future values to arrive at the future value of the annuity. This amount is $5,525.63.

We can arrive at a general annuity formula if we define the annuity amount as A, the number of time periods as N, and the interest rate per period as r. We can then define the future value as

$$\text{FV}_N = A[(1+r)^{N-1} + (1+r)^{N-2} + (1+r)^{N-3} + \ldots + (1+r)^1 + (1+r)^0]$$

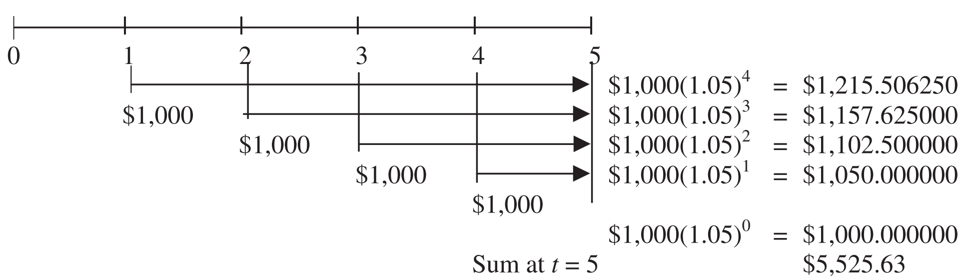
which simplifies to

$$\text{FV}_N = A\left[\frac{(1+r)^N - 1}{r}\right] \tag{7}$$

The term in brackets is the future value annuity factor. This factor gives the future value of an ordinary annuity of $1 per period. Multiplying the future value annuity factor by the annuity amount gives the future value of an ordinary annuity. For the ordinary annuity in Figure 3, we find the future value annuity factor from Equation 7 as

$$\left[\frac{(1.05)^5 - 1}{0.05}\right] = 5.525631$$

FIGURE 3 The Future Value of a Five-Year Ordinary Annuity



With an annuity amount $A = \$1{,}000$, the future value of the annuity is $\$1{,}000(5.525631) = \$5{,}525.63$, an amount that agrees with our earlier work.

The next example illustrates how to find the future value of an ordinary annuity using the formula in Equation 7.

EXAMPLE 7 The Future Value of an Annuity

Suppose your company's defined contribution retirement plan allows you to invest up to €20,000 per year. You plan to invest €20,000 per year in a stock index fund for the next 30 years. Historically, this fund has earned 9 percent per year on average. Assuming that you actually earn 9 percent a year, how much money will you have available for retirement after making the last payment?

Solution: Use Equation 7 to find the future amount:

$$A = €20{,}000$$
$$r = 9\% = 0.09$$
$$N = 30$$
$$\text{FV annuity factor} = \frac{(1+r)^N - 1}{r} = \frac{(1.09)^{30} - 1}{0.09} = 136.307539$$
$$\text{FV}_N = €20{,}000(136.307539)$$
$$= €2{,}726{,}150.77$$

Assuming the fund continues to earn an average of 9 percent per year, you will have €2,726,150.77 available at retirement.

4.2. Unequal Cash Flows

In many cases, cash flow streams are unequal, precluding the simple use of the future value annuity factor. For instance, an individual investor might have a savings plan that involves unequal cash payments depending on the month of the year or lower savings during a planned vacation. One can always find the future value of a series of unequal cash flows by compounding the cash flows one at a time. Suppose you have the five cash flows described in Table 2, indexed relative to the present ($t = 0$).

All of the payments shown in Table 2 are different. Therefore, the most direct approach to finding the future value at $t = 5$ is to compute the future value of each payment as of $t = 5$ and then sum the individual future values. The total future value at Year 5 equals \$19,190.76, as shown in the third column. Later in this chapter, you will learn shortcuts to take when the cash flows are close to even; these shortcuts will allow you to combine annuity and single-period calculations.

TABLE 2 A Series of Unequal Cash Flows and Their Future Values at 5 Percent

Time	Cash Flow ($)	Future Value at Year 5
$t = 1$	1,000	$\$1{,}000(1.05)^4 = \$1{,}215.51$
$t = 2$	2,000	$\$2{,}000(1.05)^3 = \$2{,}315.25$
$t = 3$	4,000	$\$4{,}000(1.05)^2 = \$4{,}410.00$
$t = 4$	5,000	$\$5{,}000(1.05)^1 = \$5{,}250.00$
$t = 5$	6,000	$\$6{,}000(1.05)^0 = \$6{,}000.00$
		Sum = $19,190.76

5. THE PRESENT VALUE OF A SINGLE CASH FLOW

5.1. Finding the Present Value of a Single Cash Flow

Just as the future value factor links today's present value with tomorrow's future value, the present value factor allows us to discount future value to present value. For example, with a 5 percent interest rate generating a future payoff of $105 in one year, what current amount invested at 5 percent for one year will grow to $105? The answer is $100; therefore, $100 is the present value of $105 to be received in one year at a discount rate of 5 percent.

Given a future cash flow that is to be received in N periods and an interest rate per period of r, we can use the formula for future value to solve directly for the present value as follows:

$$
\begin{aligned}
\mathrm{FV}_N &= \mathrm{PV}(1+r)^N \\
\mathrm{PV} &= \mathrm{FV}_N\left[\frac{1}{(1+r)^N}\right] \\
\mathrm{PV} &= \mathrm{FV}_N(1+r)^{-N}
\end{aligned} \tag{8}
$$

We see from Equation 8 that the present value factor, $(1 + r)^{-N}$, is the reciprocal of the future value factor, $(1 + r)^N$.

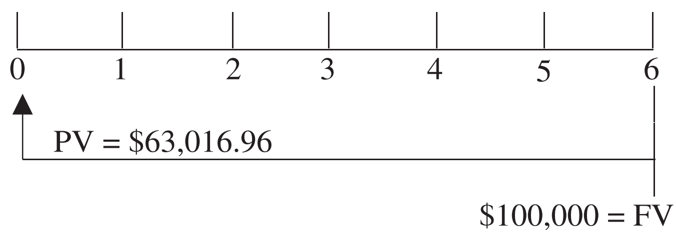
EXAMPLE 8 The Present Value of a Lump Sum

An insurance company has issued a Guaranteed Investment Contract (GIC) that promises to pay $100,000 in six years with an 8 percent return rate. What amount of money must the insurer invest today at 8 percent for six years to make the promised payment?

Solution: We can use Equation 8 to find the present value using the following data:

$$\begin{aligned}
\text{FV}_N &= \$100{,}000 \\
r &= 8\% = 0.08 \\
N &= 6 \\
\text{PV} &= \text{FV}_N(1+r)^{-N} \\
&= \$100{,}000\left[\frac{1}{(1.08)^6}\right] \\
&= \$100{,}000(0.6301696) \\
&= \$63{,}016.96
\end{aligned}$$

FIGURE 4 The Present Value of a Lump Sum to Be Received at Time $t = 6$



We can say that $63,016.96 today, with an interest rate of 8 percent, is equivalent to $100,000 to be received in six years. Discounting the $100,000 makes a future $100,000 equivalent to $63,016.96 when allowance is made for the time value of money. As the time line in Figure 4 shows, the $100,000 has been discounted six full periods.

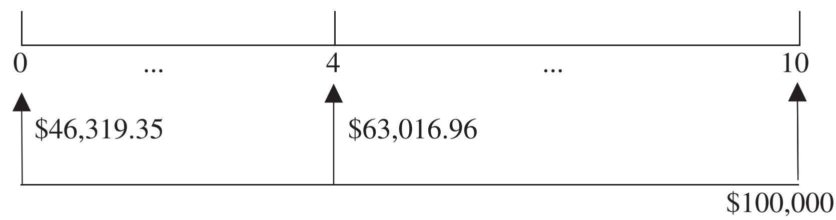
EXAMPLE 9 The Projected Present Value of a More Distant Future Lump Sum

Suppose you own a liquid financial asset that will pay you $100,000 in 10 years from today. Your daughter plans to attend college four years from today, and you want to know what the asset's present value will be at that time. Given an 8 percent discount rate, what will the asset be worth four years from today?

Solution: The value of the asset is the present value of the asset's promised payment. At $t = 4$, the cash payment will be received six years later. With this information, you can solve for the value four years from today using Equation 8:

$$
\begin{aligned}
\mathrm{FV}_N &= \$100{,}000 \\
r &= 8\% = 0.08 \\
N &= 6 \\
\mathrm{PV} &= \mathrm{FV}_N(1+r)^{-N} \\
&= \$100{,}000 \frac{1}{(1.08)^6} \\
&= \$100{,}000(0.6301696) \\
&= \$63{,}016.96
\end{aligned}
$$

FIGURE 5 The Relationship between Present Value and Future Value



The time line in Figure 5 shows the future payment of \$100,000 that is to be received at $t = 10$. The time line also shows the values at $t = 4$ and at $t = 0$. Relative to the payment at $t = 10$, the amount at $t = 4$ is a projected present value, while the amount at $t = 0$ is the present value (as of today).

Present value problems require an evaluation of the present value factor, $(1 + r)^{-N}$. Present values relate to the discount rate and the number of periods in the following ways:

- For a given discount rate, the farther in the future the amount to be received, the smaller that amount's present value.
- Holding time constant, the larger the discount rate, the smaller the present value of a future amount.

5.2. The Frequency of Compounding

Recall that interest may be paid semiannually, quarterly, monthly, or even daily. To handle interest payments made more than once a year, we can modify the present value formula (Equation 8) as follows. Recall that r_s is the quoted interest rate and equals the periodic interest rate multiplied by the number of compounding periods in each year. In general, with more than one compounding period in a year, we can express the formula for present value as

$$PV = FV_N\left(1 + \frac{r_s}{m}\right)^{-mN} \quad (9)$$

where

m = number of compounding periods per year
r_s = quoted annual interest rate
N = number of years

The formula in Equation 9 is quite similar to that in Equation 8. As we have already noted, present value and future value factors are reciprocals. Changing the frequency of compounding does not alter this result. The only difference is the use of the periodic interest rate and the corresponding number of compounding periods.

The following example illustrates Equation 9.

EXAMPLE 10 The Present Value of a Lump Sum with Monthly Compounding

The manager of a Canadian pension fund knows that the fund must make a lump-sum payment of C$5 million 10 years from now. She wants to invest an amount today in a GIC so that it will grow to the required amount. The current interest rate on GICs is 6 percent a year, compounded monthly. How much should she invest today in the GIC?

Solution: Use Equation 9 to find the required present value:

$$\begin{aligned}
FV_N &= \text{C\$5,000,000} \\
r_s &= 6\% = 0.06 \\
m &= 12 \\
r_s/m &= 0.06/12 = 0.005 \\
N &= 10 \\
mN &= 12(10) = 120 \\
PV &= FV_N\left(1 + \tfrac{r_s}{m}\right)^{-mN} \\
&= \text{C\$5,000,000}(1.005)^{-120} \\
&= \text{C\$5,000,000}(0.549633) \\
&= \text{C\$2,748,163.67}
\end{aligned}$$

In applying Equation 9, we use the periodic rate (in this case, the monthly rate) and the appropriate number of periods with monthly compounding (in this case, 10 years of monthly compounding, or 120 periods).

6. THE PRESENT VALUE OF A SERIES OF CASH FLOWS

Many applications in investment management involve assets that offer a series of cash flows over time. The cash flows may be highly uneven, relatively even, or equal. They may occur over relatively short periods of time, longer periods of time, or even stretch on indefinitely. In this section, we discuss how to find the present value of a series of cash flows.

6.1. The Present Value of a Series of Equal Cash Flows

We begin with an ordinary annuity. Recall that an ordinary annuity has equal annuity payments, with the first payment starting one period into the future. In total, the annuity makes N payments, with the first payment at $t = 1$ and the last at $t = N$. We can express the present value of an ordinary annuity as the sum of the present values of each individual annuity payment, as follows:

$$\mathrm{PV} = \frac{A}{(1+r)} + \frac{A}{(1+r)^2} + \frac{A}{(1+r)^3} + \ldots + \frac{A}{(1+r)^{N-1}} + \frac{A}{(1+r)^N} \qquad (10)$$

where

A = the annuity amount

r = the interest rate per period corresponding to the frequency of annuity payments (for example, annual, quarterly, or monthly)

N = the number of annuity payments

Because the annuity payment (A) is a constant in this equation, it can be factored out as a common term. Thus the sum of the interest factors has a shortcut expression:

$$\mathrm{PV} = A\left[\frac{1 - \frac{1}{(1+r)^N}}{r}\right] \qquad (11)$$

In much the same way that we computed the future value of an ordinary annuity, we find the present value by multiplying the annuity amount by a present value annuity factor (the term in brackets in Equation 11).

EXAMPLE 11 The Present Value of an Ordinary Annuity

Suppose you are considering purchasing a financial asset that promises to pay €1,000 per year for five years, with the first payment one year from now. The required rate of return is 12 percent per year. How much should you pay for this asset?

Solution: To find the value of the financial asset, use the formula for the present value of an ordinary annuity given in Equation 11 with the following data:

$$A = €1{,}000$$
$$r = 12\% = 0.12$$
$$N = 5$$

$$\mathrm{PV} = A\left[\frac{1-\frac{1}{(1+r)^N}}{r}\right]$$
$$= €1{,}000\left[\frac{1-\frac{1}{(1.12)^5}}{0.12}\right]$$
$$= €1{,}000(3.604776)$$
$$= €3{,}604.78$$

The series of cash flows of €1,000 per year for five years is currently worth €3,604.78 when discounted at 12 percent.

FIGURE 6 An Annuity Due of $100 per Period

Keeping track of the actual calendar time brings us to a specific type of annuity with level payments: the annuity due. An annuity due has its first payment occurring today ($t = 0$). In total, the annuity due will make N payments. Figure 6 presents the time line for an annuity due that makes four payments of $100.

As Figure 6 shows, we can view the four-period annuity due as the sum of two parts: a $100 lump sum today and an ordinary annuity of $100 per period for three periods. At a 12 percent discount rate, the four $100 cash flows in this annuity due example will be worth $340.18.[7]

Expressing the value of the future series of cash flows in today's dollars gives us a convenient way of comparing annuities. The next example illustrates this approach.

EXAMPLE 12 An Annuity Due as the Present Value of an Immediate Cash Flow Plus an Ordinary Annuity

You are retiring today and must choose to take your retirement benefits either as a lump sum or as an annuity. Your company's benefits officer presents you with two

[7]There is an alternative way to calculate the present value of an annuity due. Compared to an ordinary annuity, the payments in an annuity due are each discounted one less period. Therefore, we can modify Equation 11 to handle annuities due by multiplying the right-hand side of the equation by $(1 + r)$:

$$\mathrm{PV(Annuity\ due)} = A\{[1 - (1 + r)^{-N}]/r\}(1 + r)$$

alternatives: an immediate lump sum of \$2 million or an annuity with 20 payments of \$200,000 a year with the first payment starting today. The interest rate at your bank is 7 percent per year compounded annually. Which option has the greater present value? (Ignore any tax differences between the two options.)

Solution: To compare the two options, find the present value of each at time $t = 0$ and choose the one with the larger value. The first option's present value is \$2 million, already expressed in today's dollars. The second option is an annuity due. Because the first payment occurs at $t = 0$, you can separate the annuity benefits into two pieces: an immediate \$200,000 to be paid today ($t = 0$) and an ordinary annuity of \$200,000 per year for 19 years. To value this option, you need to find the present value of the ordinary annuity using Equation 11 and then add \$200,000 to it.

$$\begin{aligned}
A &= \$200{,}000 \\
N &= 19 \\
r &= 7\% = 0.07 \\
\text{PV} &= A\left[\frac{1-\dfrac{1}{(1+r)^N}}{r}\right] \\
&= \$200{,}000\left[\frac{1-\dfrac{1}{(1.07)^{19}}}{0.07}\right] \\
&= \$200{,}000(10.335595) \\
&= \$2{,}067{,}119.05
\end{aligned}$$

The 19 payments of \$200,000 have a present value of \$2,067,119.05. Adding the initial payment of \$200,000 to \$2,067,119.05, we find that the total value of the annuity option is \$2,267,119.05. The present value of the annuity is greater than the lump sum alternative of \$2 million.

We now look at another example reiterating the equivalence of present and future values.

EXAMPLE 13 The Projected Present Value of an Ordinary Annuity

A German pension fund manager anticipates that benefits of €1 million per year must be paid to retirees. Retirements will not occur until 10 years from now at time $t = 10$. Once benefits begin to be paid, they will extend until $t = 39$ for a total of 30 payments. What is the present value of the pension liability if the appropriate annual discount rate for plan liabilities is 5 percent compounded annually?

Solution: This problem involves an annuity with the first payment at $t = 10$. From the perspective of $t = 9$, we have an ordinary annuity with 30 payments. We can compute the present value of this annuity with Equation 11 and then look at it on a time line.

$$\begin{aligned} A &= €1{,}000{,}000 \\ r &= 5\% = 0.05 \\ N &= 30 \\ \text{PV} &= A\left[\frac{1-\frac{1}{(1+r)^N}}{r}\right] \\ &= €1{,}000{,}000\left[\frac{1-\frac{1}{(1.05)^{30}}}{0.05}\right] \\ &= €1{,}000{,}000(15.372451) \\ &= €15{,}372{,}451.03 \end{aligned}$$

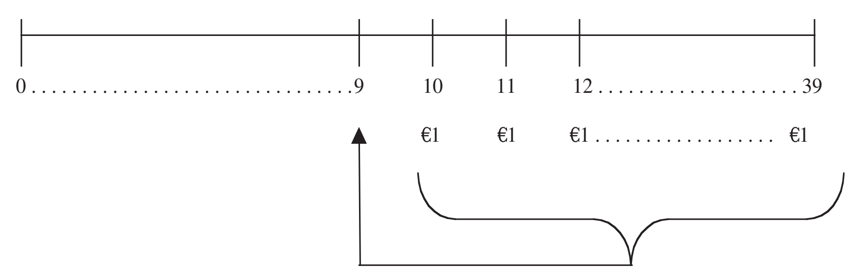
On the time line, we have shown the pension payments of €1 million extending from $t = 10$ to $t = 39$. The bracket and arrow indicate the process of finding the present value of the annuity, discounted back to $t = 9$. The present value of the pension benefits as of $t = 9$ is €15,372,451.03. The problem is to find the present value today (at $t = 0$).

Now we can rely on the equivalence of present value and future value. As Figure 7 shows, we can view the amount at $t = 9$ as a future value from the vantage point of $t = 0$. We compute the present value of the amount at $t = 9$ as follows:

$$\begin{aligned} \text{FV}_N &= €15{,}372{,}451.03 \text{ (the present value at } t = 9) \\ N &= 9 \\ r &= 5\% = 0.05 \\ \text{PV} &= \text{FV}_N(1 + r)^{-N} \\ &= €15{,}372{,}451.03(1.05)^{-9} \\ &= €15{,}372{,}451.03(0.644609) \\ &= €9{,}909{,}219.00 \end{aligned}$$

The present value of the pension liability is €9,909,219.00.

FIGURE 7 The Present Value of an Ordinary Annuity with First Payment at Time $t = 10$ (in Millions)



Example 13 illustrates three procedures emphasized in this chapter:

1. finding the present or future value of any cash flow series;
2. recognizing the equivalence of present value and appropriately discounted future value; and
3. keeping track of the actual calendar time in a problem involving the time value of money.

6.2. The Present Value of an Infinite Series of Equal Cash Flows—Perpetuity

Consider the case of an ordinary annuity that extends indefinitely. Such an ordinary annuity is called a perpetuity (a perpetual annuity). To derive a formula for the present value of a perpetuity, we can modify Equation 10 to account for an infinite series of cash flows:

$$\mathrm{PV} = A\sum_{t=1}^{\infty}\left[\frac{1}{(1+r)^t}\right] \tag{12}$$

As long as interest rates are positive, the sum of present value factors converges and

$$\mathrm{PV} = \frac{A}{r} \tag{13}$$

To see this, look back at Equation 11, the expression for the present value of an ordinary annuity. As N (the number of periods in the annuity) goes to infinity, the term $1/(1 + r)^N$ approaches 0 and Equation 11 simplifies to Equation 13. This equation will reappear when we value dividends from stocks because stocks have no predefined life span. (A stock paying constant dividends is similar to a perpetuity.) With the first payment a year from now, a perpetuity of \$10 per year with a 20 percent required rate of return has a present value of \$10/0.2 = \$50.

Equation 13 is valid only for a perpetuity with level payments. In our development above, the first payment occurred at $t = 1$; therefore, we compute the present value as of $t = 0$.

Other assets also come close to satisfying the assumptions of a perpetuity. Certain government bonds and preferred stocks are typical examples of financial assets that make level payments for an indefinite period of time.

EXAMPLE 14 The Present Value of a Perpetuity

The British government once issued a type of security called a consol bond, which promised to pay a level cash flow indefinitely. If a consol bond paid £100 per year in perpetuity, what would it be worth today if the required rate of return were 5 percent?

Solution: To answer this question, we can use Equation 13 with the following data:

$$\begin{aligned} A &= £100 \\ r &= 5\% = 0.05 \\ \mathrm{PV} &= A/r \end{aligned}$$

$$= £100/0.05$$
$$= £2{,}000$$

The bond would be worth £2,000.

6.3. Present Values Indexed at Times Other than $t = 0$

In practice with investments, analysts frequently need to find present values indexed at times other than $t = 0$. Subscripting the present value and evaluating a perpetuity beginning with \$100 payments in Year 2, we find PV_1 = \$100/0.05 = \$2,000 at a 5 percent discount rate. Further, we can calculate today's PV as PV_0 = \$2,000/1.05 = \$1,904.76.

Consider a similar situation in which cash flows of \$6 per year begin at the end of the 4th year and continue at the end of each year thereafter, with the last cash flow at the end of the 10th year. From the perspective of the end of the third year, we are facing a typical seven-year ordinary annuity. We can find the present value of the annuity from the perspective of the end of the third year and then discount that present value back to the present. At an interest rate of 5 percent, the cash flows of \$6 per year starting at the end of the fourth year will be worth \$34.72 at the end of the third year ($t = 3$) and \$29.99 today ($t = 0$).

The next example illustrates the important concept that an annuity or perpetuity beginning sometime in the future can be expressed in present value terms one period prior to the first payment. That present value can then be discounted back to today's present value.

EXAMPLE 15 The Present Value of a Projected Perpetuity

Consider a level perpetuity of £100 per year with its first payment beginning at $t = 5$. What is its present value today (at $t = 0$), given a 5 percent discount rate?

Solution: First, we find the present value of the perpetuity at $t = 4$ and then discount that amount back to $t = 0$. (Recall that a perpetuity or an ordinary annuity has its first payment one period away, explaining the $t = 4$ index for our present value calculation.)

i. Find the present value of the perpetuity at $t = 4$:

$$A = £100$$
$$r = 5\% = 0.05$$
$$\text{PV} = A/r$$
$$= £100/0.05$$
$$= £2{,}000$$

ii. Find the present value of the future amount at $t = 4$. From the perspective of $t = 0$, the present value of £2,000 can be considered a future value. Now we need to find the present value of a lump sum:

$$\begin{aligned}
\text{FV}_N &= £2{,}000 \text{ (the present value at } t = 4) \\
r &= 5\% = 0.05 \\
N &= 4 \\
\text{PV} &= \text{FV}_N(1 + r)^{-N} \\
&= £2{,}000(1.05)^{-4} \\
&= £2{,}000(0.822702) \\
&= £1{,}645.40
\end{aligned}$$

Today's present value of the perpetuity is £1,645.40.

As discussed earlier, an annuity is a series of payments of a fixed amount for a specified number of periods. Suppose we own a perpetuity. At the same time, we issue a perpetuity obligating us to make payments; these payments are the same size as those of the perpetuity we own. However, the first payment of the perpetuity we issue is at $t = 5$; payments then continue on forever. The payments on this second perpetuity exactly offset the payments received from the perpetuity we own at $t = 5$ and all subsequent dates. We are left with level nonzero net cash flows at $t = 1, 2, 3$, and 4. This outcome exactly fits the definition of an annuity with four payments. Thus we can construct an annuity as the difference between two perpetuities with equal, level payments but differing starting dates. The next example illustrates this result.

EXAMPLE 16 The Present Value of an Ordinary Annuity as the Present Value of a Current Minus Projected Perpetuity

Given a 5 percent discount rate, find the present value of a four-year ordinary annuity of £100 per year starting in Year 1 as the difference between the following two level perpetuities:

Perpetuity 1	£100 per year starting in Year 1 (first payment at $t = 1$)
Perpetuity 2	£100 per year starting in Year 5 (first payment at $t = 5$)

Solution: If we subtract Perpetuity 2 from Perpetuity 1, we are left with an ordinary annuity of £100 per period for four years (payments at $t = 1, 2, 3, 4$). Subtracting the present value of Perpetuity 2 from that of Perpetuity 1, we arrive at the present value of the four-year ordinary annuity:

$$\begin{aligned}
\text{PV}_0(\text{Perpetuity 1}) &= £100/0.05 = £2{,}000 \\
\text{PV}_4(\text{Perpetuity 2}) &= £100/0.05 = £2{,}000 \\
\text{PV}_0(\text{Perpetuity 2}) &= £2{,}000/(1.05)^4 = £1{,}645.40 \\
\text{PV}_0(\text{Annuity}) &= \text{PV}_0(\text{Perpetuity 1}) - \text{PV}_0(\text{Perpetuity 2}) \\
&= £2{,}000 - £1{,}645.40 \\
&= £354.60
\end{aligned}$$

The four-year ordinary annuity's present value is equal to £2,000 – £1,645.40 = £354.60.

TABLE 3 A Series of Unequal Cash Flows and Their Present Values at 5 Percent

Time Period	**Cash Flow ($)**	**Present Value at Year 0**
1	1,000	$\$1{,}000(1.05)^{-1} = \$952.38$
2	2,000	$\$2{,}000(1.05)^{-2} = \$1{,}814.06$
3	4,000	$\$4{,}000(1.05)^{-3} = \$3{,}455.35$
4	5,000	$\$5{,}000(1.05)^{-4} = \$4{,}113.51$
5	6,000	$\$6{,}000(1.05)^{-5} = \$4{,}701.16$
		Sum = $15,036.46

6.4. The Present Value of a Series of Unequal Cash Flows

When we have unequal cash flows, we must first find the present value of each individual cash flow and then sum the respective present values. For a series with many cash flows, we usually use a spreadsheet. Table 3 lists a series of cash flows with the time periods in the first column, cash flows in the second column, and each cash flow's present value in the third column. The last row of Table 3 shows the sum of the five present values.

We could calculate the future value of these cash flows by computing them one at a time using the single-payment future value formula. We already know the present value of this series, however, so we can easily apply time-value equivalence. The future value of the series of cash flows from Table 2, $19,190.76, is equal to the single $15,036.46 amount compounded forward to $t = 5$:

$$
\begin{aligned}
\mathrm{PV} &= \$15{,}036.46 \\
N &= 5 \\
r &= 5\% = 0.05 \\
\mathrm{FV}_N &= \mathrm{PV}(1 + r)^N \\
&= \$15{,}036.46(1.05)^5 \\
&= \$15{,}036.46(1.276282) \\
&= \$19{,}190.76
\end{aligned}
$$

7. SOLVING FOR RATES, NUMBER OF PERIODS, OR SIZE OF ANNUITY PAYMENTS

In the previous examples, certain pieces of information have been made available. For instance, all problems have given the rate of interest, r, the number of time periods, N, the annuity amount, A, and either the present value, PV, or future value, FV. In real-world applications, however, although the present and future values may be given, you may have to solve for either the interest rate, the number of periods, or the annuity amount. In the subsections that follow, we show these types of problems.

7.1. Solving for Interest Rates and Growth Rates

Suppose a bank deposit of €100 is known to generate a payoff of €111 in one year. With this information, we can infer the interest rate that separates the present value of €100 from the future value of €111 by using Equation 2, $FV_N = PV(1 + r)^N$, with $N = 1$. With PV, FV, and N known, we can solve for r directly:

$$1 + r = FV/PV$$
$$1 + r = €111/€100 = 1.11$$
$$r = 0.11, \text{ or } 11\%$$

The interest rate that equates €100 at $t = 0$ to €111 at $t = 1$ is 11 percent. Thus we can state that €100 grows to €111 with a growth rate of 11 percent.

As this example shows, an interest rate can also be considered a growth rate. The particular application will usually dictate whether we use the term "interest rate" or "growth rate." Solving Equation 2 for r and replacing the interest rate r with the growth rate g produces the following expression for determining growth rates:

$$g = (FV_N/PV)^{1/N} - 1 \tag{14}$$

Below are two examples that use the concept of a growth rate.

EXAMPLE 17 Calculating a Growth Rate (1)

Hyundai Steel, the first Korean steelmaker, was established in 1953. Hyundai Steel's sales increased from ₩ 14,146.4 billion in 2012 to ₩ 19,166.0 billion in 2017. However, its net profit declined from ₩ 796.4 billion in 2012 to ₩ 727.5 billion in 2017. Calculate the following growth rates for Hyundai Steel for the five-year period from the end of 2012 to the end of 2017:

1. Sales growth rate
2. Net profit growth rate

Solution to 1: To solve this problem, we can use Equation 14, $g = (FV_N/PV)^{1/N} - 1$. We denote sales in 2012 as PV and sales in 2017 as FV_5. We can then solve for the growth rate as follows:

$$\begin{aligned} g &= \sqrt[5]{₩\,19{,}166.0/₩\,14{,}146.4} - 1 \\ &= \sqrt[5]{1.354832} - 1 \\ &= 1.062618 - 1 \\ &= 0.062618 \text{ or about } 6.3\% \end{aligned}$$

The calculated growth rate of about 6.3 percent a year shows that Hyundai Steel's sales grew during the 2012–2017 period.

Solution to 2: In this case, we can speak of a positive compound rate of decrease or a negative compound growth rate. Using Equation 14, we find

$$\begin{aligned} g &= \sqrt[5]{₩727.5/₩796.4} - 1 \\ &= \sqrt[5]{0.913486} - 1 \\ &= 0.982065 - 1 \\ &= -0.017935 \text{ or about } -1.8\% \end{aligned}$$

In contrast to the positive sales growth, the rate of growth in net profit was approximately –1.8 percent during the 2012–2017 period.

EXAMPLE 18 Calculating a Growth Rate (2)

Toyota Motor Corporation, one of the largest automakers in the world, had consolidated vehicle sales of 8.96 million units in 2018 (fiscal year ending 31 March 2018). This is substantially more than consolidated vehicle sales of 7.35 million units six years earlier in 2012. What was the growth rate in number of vehicles sold by Toyota from 2012 to 2018?

Solution: Using Equation 14, we find

$$\begin{aligned} g &= \sqrt[6]{8.96/7.35} - 1 \\ &= \sqrt[6]{1.219048} - 1 \\ &= 1.033563 - 1 \\ &= 0.033563 \text{ or about } 3.4\% \end{aligned}$$

The rate of growth in vehicles sold was approximately 3.4 percent during the 2012–2018 period. Note that we can also refer to 3.4 percent as the compound annual growth rate because it is the single number that compounds the number of vehicles sold in 2012 forward to the number of vehicles sold in 2018. Table 4 lists the number of vehicles sold by Toyota from 2012 to 2018.

Table 4 also shows 1 plus the one-year growth rate in number of vehicles sold. We can compute the 1 plus six-year cumulative growth in number of vehicles sold from 2012 to 2018 as the product of quantities (1 + one-year growth rate). We arrive at the same result as when we divide the ending number of vehicles sold, 8.96 million, by the beginning number of vehicles sold, 7.35 million:

$$\frac{8.96}{7.35} = \left(\frac{8.87}{7.35}\right)\left(\frac{9.12}{8.87}\right)\left(\frac{8.97}{9.12}\right)\left(\frac{8.68}{8.97}\right)\left(\frac{8.97}{8.68}\right)\left(\frac{8.96}{8.97}\right)$$

$$= (1 + g_1)(1 + g_2)(1 + g_3)(1 + g_4)(1 + g_5)(1 + g_6)$$

$$1.219048 = (1.206803)(1.028185)(0.983553)(0.967670)(1.033410)(0.998885)$$

TABLE 4 Number of Vehicles Sold, 2012–2018

Year	Number of Vehicles Sold (Millions)	$(1 + g)_t$	t
2012	7.35		0
2013	8.87	8.87/7.35 = 1.206803	1
2014	9.12	9.12/8.87 = 1.028185	2
2015	8.97	8.97/9.12 = 0.983553	3
2016	8.68	8.68/8.97 = 0.967670	4
2017	8.97	8.97/8.68 = 1.033410	5
2018	8.96	8.96/8.97 = 0.998885	6

Source: www.toyota.com.

The right-hand side of the equation is the product of 1 plus the one-year growth rate in number of vehicles sold for each year. Recall that, using Equation 14, we took the sixth root of 8.96/7.35 = 1.219048. In effect, we were solving for the single value of g which, when compounded over six periods, gives the correct product of 1 plus the one-year growth rates.[8]

In conclusion, we do not need to compute intermediate growth rates as in Table 4 to solve for a compound growth rate g. Sometimes, however, the intermediate growth rates are interesting or informative. For example, most of the 21.9 percent increase in vehicles sold by Toyota from 2012 to 2018 occurred in 2013 as sales increased by 20.7 percent from 2012 to 2013. Elsewhere in Toyota Motor's disclosures, the company noted that all regions except Europe showed a substantial increase in sales in 2013. We can also analyze the variability in growth rates when we conduct an analysis as in Table 4. Sales continued to increase in 2014 but then declined in 2015 and 2016. Sales then increased but the sales in 2017 and 2018 are about the same as in 2015.

The compound growth rate is an excellent summary measure of growth over multiple time periods. In our Toyota Motors example, the compound growth rate of 3.4 percent is the single growth rate that, when added to 1, compounded over six years, and multiplied by the 2012 number of vehicles sold, yields the 2018 number of vehicles sold.

7.2. Solving for the Number of Periods

In this section, we demonstrate how to solve for the number of periods given present value, future value, and interest or growth rates.

[8]The compound growth rate that we calculate here is an example of a geometric mean, specifically the geometric mean of the growth rates. We define the geometric mean in Chapter 2, which is on statistical concepts.

EXAMPLE 19 The Number of Annual Compounding Periods Needed for an Investment to Reach a Specific Value

You are interested in determining how long it will take an investment of €10,000,000 to double in value. The current interest rate is 7 percent compounded annually. How many years will it take €10,000,000 to double to €20,000,000?

Solution: Use Equation 2, $FV_N = PV(1 + r)^N$, to solve for the number of periods, N, as follows:

$$\begin{aligned}(1+r)^N &= FV_N/PV = 2\\ N\ln(1+r) &= \ln(2)\\ N &= \ln(2)/\ln(1+r)\\ &= \ln(2)/\ln(1.07) = 10.24\end{aligned}$$

With an interest rate of 7 percent, it will take approximately 10 years for the initial €10,000,000 investment to grow to €20,000,000. Solving for N in the expression $(1.07)^N = 2.0$ requires taking the natural logarithm of both sides and using the rule that $\ln(x^N) = N\ln(x)$. Generally, we find that $N = [\ln(FV/PV)]/\ln(1 + r)$. Here, $N = \ln(€20{,}000{,}000/€10{,}000{,}000)/\ln(1.07) = \ln(2)/\ln(1.07) = 10.24$.[9]

7.3. Solving for the Size of Annuity Payments

In this section, we discuss how to solve for annuity payments. Mortgages, auto loans, and retirement savings plans are classic examples of applications of annuity formulas.

EXAMPLE 20 Calculating the Size of Payments on a Fixed-Rate Mortgage

You are planning to purchase a $120,000 house by making a down payment of $20,000 and borrowing the remainder with a 30-year fixed-rate mortgage with monthly payments. The first payment is due at $t = 1$. Current mortgage interest rates are quoted at 8 percent with monthly compounding. What will your monthly mortgage payments be?

[9]To quickly approximate the number of periods, practitioners sometimes use an ad hoc rule called the **Rule of 72**: Divide 72 by the stated interest rate to get the approximate number of years it would take to double an investment at the interest rate. Here, the approximation gives 72/7 = 10.3 years. The Rule of 72 is loosely based on the observation that it takes 12 years to double an amount at a 6 percent interest rate, giving $6 \times 12 = 72$. At a 3 percent rate, one would guess it would take twice as many years, $3 \times 24 = 72$.

Solution: The bank will determine the mortgage payments such that at the stated periodic interest rate, the present value of the payments will be equal to the amount borrowed (in this case, \$100,000). With this fact in mind, we can use Equation 11, $PV = A\left[\frac{1-\frac{1}{(1+r)^N}}{r}\right]$, to solve for the annuity amount, A, as the present value divided by the present value annuity factor:

$$
\begin{aligned}
PV &= \$100{,}000 \\
r_s &= 8\% = 0.08 \\
m &= 12 \\
r_s/m &= 0.08/12 = 0.006667 \\
N &= 30 \\
mN &= 12 \times 30 = 360
\end{aligned}
$$

$$
\text{Present value annuity factor} = \frac{1 - \frac{1}{[1+(r_s/m)]^{mN}}}{r_s/m} = \frac{1 - \frac{1}{(1.006667)^{360}}}{0.006667}
$$

$$
\begin{aligned}
&= 136.283494 \\
A &= PV/\text{Present value annuity factor} \\
&= \$100{,}000/136.283494 \\
&= \$733.76
\end{aligned}
$$

The amount borrowed, \$100,000, is equivalent to 360 monthly payments of \$733.76 with a stated interest rate of 8 percent. The mortgage problem is a relatively straightforward application of finding a level annuity payment.

Next, we turn to a retirement-planning problem. This problem illustrates the complexity of the situation in which an individual wants to retire with a specified retirement income. Over the course of a life cycle, the individual may be able to save only a small amount during the early years but then may have the financial resources to save more during later years. Savings plans often involve uneven cash flows, a topic we will examine in the last part of this chapter. When dealing with uneven cash flows, we take maximum advantage of the principle that dollar amounts indexed at the same point in time are additive—the **cash flow additivity principle**.

EXAMPLE 21 The Projected Annuity Amount Needed to Fund a Future-Annuity Inflow

Jill Grant is 22 years old (at $t = 0$) and is planning for her retirement at age 63 (at $t = 41$). She plans to save \$2,000 per year for the next 15 years ($t = 1$ to $t = 15$). She wants to have retirement income of \$100,000 per year for 20 years, with the first retirement payment starting at $t = 41$. How much must Grant save each year from $t = 16$ to $t = 40$

FIGURE 8 Solving for Missing Annuity Payments (in Thousands)

0	1	2		15	16	17		40	41	42		60
	(\$2)	(\$2)	...	(\$2)	(X)	(X)	...	(X)	\$100	\$100	...	\$100

in order to achieve her retirement goal? Assume she plans to invest in a diversified stock-and-bond mutual fund that will earn 8 percent per year on average.

Solution: To help solve this problem, we set up the information on a time line. As Figure 8 shows, Grant will save \$2,000 (an outflow) each year for Years 1 to 15. Starting in Year 41, Grant will start to draw retirement income of \$100,000 per year for 20 years. In the time line, the annual savings is recorded in parentheses (\$2) to show that it is an outflow. The problem is to find the savings, recorded as X, from Year 16 to Year 40.

Solving this problem involves satisfying the following relationship: The present value of savings (outflows) equals the present value of retirement income (inflows). We could bring all the dollar amounts to $t = 40$ or to $t = 15$ and solve for X.

Let us evaluate all dollar amounts at $t = 15$ (we encourage the reader to repeat the problem by bringing all cash flows to $t = 40$). As of $t = 15$, the first payment of X will be one period away (at $t = 16$). Thus we can value the stream of Xs using the formula for the present value of an ordinary annuity.

This problem involves three series of level cash flows. The basic idea is that the present value of the retirement income must equal the present value of Grant's savings. Our strategy requires the following steps:

1. Find the future value of the savings of \$2,000 per year and index it at $t = 15$. This value tells us how much Grant will have saved.
2. Find the present value of the retirement income at $t = 15$. This value tells us how much Grant needs to meet her retirement goals (as of $t = 15$). Two substeps are necessary. First, calculate the present value of the annuity of \$100,000 per year at $t = 40$. Use the formula for the present value of an annuity. (Note that the present value is indexed at $t = 40$ because the first payment is at $t = 41$.) Next, discount the present value back to $t = 15$ (a total of 25 periods).
3. Now compute the difference between the amount Grant has saved (Step 1) and the amount she needs to meet her retirement goals (Step 2). Her savings from $t = 16$ to $t = 40$ must have a present value equal to the difference between the future value of her savings and the present value of her retirement income.

Our goal is to determine the amount Grant should save in each of the 25 years from $t = 16$ to $t = 40$. We start by bringing the \$2,000 savings to $t = 15$, as follows:

$$A = \$2{,}000$$
$$r = 8\% = 0.08$$
$$N = 15$$
$$\text{FV} = A\left[\frac{(1+r)^N - 1}{r}\right]$$
$$= \$2{,}000\left[\frac{(1.08)^{15} - 1}{0.08}\right]$$
$$= \$2{,}000(27.152114)$$
$$= \$54{,}304.23$$

At $t = 15$, Grant's initial savings will have grown to \$54,304.23.

Now we need to know the value of Grant's retirement income at $t = 15$. As stated earlier, computing the retirement present value requires two substeps. First, find the present value at $t = 40$ with the formula in Equation 11; second, discount this present value back to $t = 15$. Now we can find the retirement income present value at $t = 40$:

$$A = \$100{,}000$$
$$r = 8\% = 0.08$$
$$N = 20$$
$$\text{PV} = A\left[\frac{1 - \dfrac{1}{(1+r)^N}}{r}\right]$$
$$= \$100{,}000\left[\frac{1 - \dfrac{1}{(1.08)^{20}}}{0.08}\right]$$
$$= \$100{,}000(9.818147)$$
$$= \$981{,}814.74$$

The present value amount is as of $t = 40$, so we must now discount it back as a lump sum to $t = 15$:

$$\text{FV}_N = \$981{,}814.74$$
$$N = 25$$
$$r = 8\% = 0.08$$
$$\text{PV} = \text{FV}_N(1+r)^{-N}$$
$$= \$981{,}814.74(1.08)^{-25}$$
$$= \$981{,}814.74(0.146018)$$
$$= \$143{,}362.53$$

Now recall that Grant will have saved \$54,304.23 by $t = 15$. Therefore, in present value terms, the annuity from $t = 16$ to $t = 40$ must equal the difference between the amount already saved (\$54,304.23) and the amount required for retirement (\$143,362.53). This amount is equal to \$143,362.53 – \$54,304.23 = \$89,058.30. Therefore, we must now find the annuity payment, A, from $t = 16$ to $t = 40$ that has a present value of \$89,058.30. We find the annuity payment as follows:

$$
\begin{aligned}
\text{PV} &= \$89{,}058.30 \\
r &= 8\% = 0.08 \\
N &= 25 \\
\text{Present value annuity factor} &= \left[\frac{1 - \dfrac{1}{(1+r)^N}}{r}\right] \\
&= \left[\frac{1 - \dfrac{1}{(1.08)^{25}}}{0.08}\right] \\
&= 10.674776 \\
A &= \text{PV/Present value annuity factor} \\
&= \$89{,}058.30/10.674776 \\
&= \$8{,}342.87
\end{aligned}
$$

Grant will need to increase her savings to \$8,342.87 per year from $t = 16$ to $t = 40$ to meet her retirement goal of having a fund equal to \$981,814.74 after making her last payment at $t = 40$.

7.4. Review of Present and Future Value Equivalence

As we have demonstrated, finding present and future values involves moving amounts of money to different points on a time line. These operations are possible because present value and future value are equivalent measures separated in time. Table 5 illustrates this equivalence; it lists the timing of five cash flows, their present values at $t = 0$, and their future values at $t = 5$.

To interpret Table 5, start with the third column, which shows the present values. Note that each \$1,000 cash payment is discounted back the appropriate number of periods to find the present value at $t = 0$. The present value of \$4,329.48 is exactly equivalent to the series of cash flows. This information illustrates an important point: A lump sum can actually generate an annuity. If we place a lump sum in an account that earns the stated interest rate for all periods, we can generate an annuity that is equivalent to the lump sum. Amortized loans, such as mortgages and car loans, are examples of this principle.

TABLE 5 The Equivalence of Present and Future Values

Time	Cash Flow ($)	Present Value at $t = 0$	Future Value at $t = 5$
1	1,000	$\$1{,}000(1.05)^{-1} = \$952.38$	$\$1{,}000(1.05)^{4} = \$1{,}215.51$
2	1,000	$\$1{,}000(1.05)^{-2} = \$907.03$	$\$1{,}000(1.05)^{3} = \$1{,}157.63$
3	1,000	$\$1{,}000(1.05)^{-3} = \$863.84$	$\$1{,}000(1.05)^{2} = \$1{,}102.50$
4	1,000	$\$1{,}000(1.05)^{-4} = \$822.70$	$\$1{,}000(1.05)^{1} = \$1{,}050.00$
5	1,000	$\$1{,}000(1.05)^{-5} = \$783.53$	$\$1{,}000(1.05)^{0} = \$1{,}000.00$
		Sum: $4,329.48	Sum: $5,525.64

To see how a lump sum can fund an annuity, assume that we place $4,329.48 in the bank today at 5 percent interest. We can calculate the size of the annuity payments by using Equation 11. Solving for A, we find

$$\begin{aligned} A &= \frac{\text{PV}}{\dfrac{1 - [1/(1+r)^N]}{r}} \\ &= \frac{\$4{,}329.48}{\dfrac{1 - [1/(1.05)^5]}{0.05}} \\ &= \$1{,}000 \end{aligned}$$

Table 6 shows how the initial investment of $4,329.48 can actually generate five $1,000 withdrawals over the next five years.

To interpret Table 6, start with an initial present value of $4,329.48 at $t = 0$. From $t = 0$ to $t = 1$, the initial investment earns 5 percent interest, generating a future value of $4,329.48(1.05) = $4,545.95. We then withdraw $1,000 from our account, leaving $4,545.95 – $1,000 = $3,545.95 (the figure reported in the last column for time period 1). In the next period, we earn one year's worth of interest and then make a $1,000 withdrawal. After the fourth withdrawal, we have $952.38, which earns 5 percent. This amount then grows to $1,000 during the year, just enough for us to make the last withdrawal. Thus the

TABLE 6 How an Initial Present Value Funds an Annuity

Time Period	Amount Available at the Beginning of the Time Period ($)	Ending Amount before Withdrawal	Withdrawal ($)	Amount Available after Withdrawal ($)
1	4,329.48	$4,329.48(1.05) = $4,545.95	1,000	3,545.95
2	3,545.95	$3,545.95(1.05) = $3,723.25	1,000	2,723.25
3	2,723.25	$2,723.25(1.05) = $2,859.41	1,000	1,859.41
4	1,859.41	$1,859.41(1.05) = $1,952.38	1,000	952.38
5	952.38	$952.38(1.05) = $1,000	1,000	0

initial present value, when invested at 5 percent for five years, generates the $1,000 five-year ordinary annuity. The present value of the initial investment is exactly equivalent to the annuity.

Now we can look at how future value relates to annuities. In Table 5, we reported that the future value of the annuity was $5,525.64. We arrived at this figure by compounding the first $1,000 payment forward four periods, the second $1,000 forward three periods, and so on. We then added the five future amounts at $t = 5$. The annuity is equivalent to $5,525.64 at $t = 5$ and $4,329.48 at $t = 0$. These two dollar measures are thus equivalent. We can verify the equivalence by finding the present value of $5,525.64, which is $\$5{,}525.64 \times (1.05)^{-5} = \$4{,}329.48$. We found this result above when we showed that a lump sum can generate an annuity.

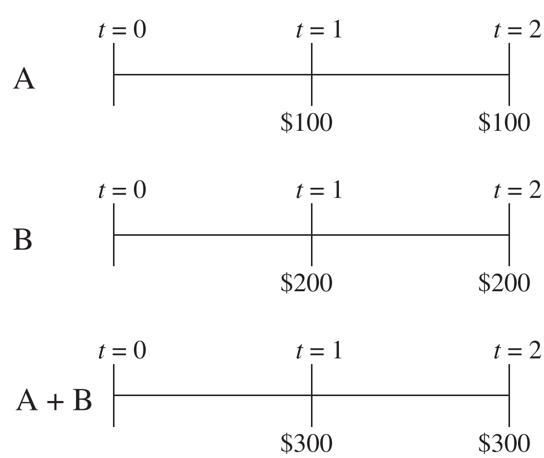
To summarize what we have learned so far: A lump sum can be seen as equivalent to an annuity, and an annuity can be seen as equivalent to its future value. Thus present values, future values, and a series of cash flows can all be considered equivalent as long as they are indexed at the same point in time.

7.5. The Cash Flow Additivity Principle

The cash flow additivity principle—the idea that amounts of money indexed at the same point in time are additive—is one of the most important concepts in time value of money mathematics. We have already mentioned and used this principle; this section provides a reference example for it.

Consider the two series of cash flows shown on the time line in Figure 9. The series are denoted A and B. If we assume that the annual interest rate is 2 percent, we can find the future value of each series of cash flows as follows. Series A's future value is $\$100(1.02) + \$100 = \$202$. Series B's future value is $\$200(1.02) + \$200 = \$404$. The future value of (A + B) is $\$202 + \$404 = \$606$ by the method we have used up to this point. The alternative way to find the future value is to add the cash flows of each series, A and B (call it A + B), and then find the future value of the combined cash flow, as shown in Figure 9.

FIGURE 9 The Additivity of Two Series of Cash Flows



The third time line in Figure 9 shows the combined series of cash flows. Series A has a cash flow of \$100 at $t = 1$, and Series B has a cash flow of \$200 at $t = 1$. The combined series thus has a cash flow of \$300 at $t = 1$. We can similarly calculate the cash flow of the combined series at $t = 2$. The future value of the combined series (A + B) is \$300(1.02) + \$300 = \$606, the same result we found when we added the future values of each series.

The additivity and equivalence principles also appear in another common situation. Suppose cash flows are \$4 at the end of the first year and \$24 (actually separate payments of \$4 and \$20) at the end of the second year. Rather than finding present values of the first year's \$4 and the second year's \$24, we can treat this situation as a \$4 annuity for two years and a second-year \$20 lump sum. If the discount rate were 6 percent, the \$4 annuity would have a present value of \$7.33 and the \$20 lump sum a present value of \$17.80, for a total of \$25.13.

8. SUMMARY

In this chapter, we have explored a foundation topic in investment mathematics, the time value of money. We have developed and reviewed the following concepts for use in financial applications:

- The interest rate, r, is the required rate of return; r is also called the discount rate or opportunity cost.
- An interest rate can be viewed as the sum of the real risk-free interest rate and a set of premiums that compensate lenders for risk: an inflation premium, a default risk premium, a liquidity premium, and a maturity premium.
- The future value, FV, is the present value, PV, times the future value factor, $(1 + r)^N$.
- The interest rate, r, makes current and future currency amounts equivalent based on their time value.
- The stated annual interest rate is a quoted interest rate that does not account for compounding within the year.
- The periodic rate is the quoted interest rate per period; it equals the stated annual interest rate divided by the number of compounding periods per year.
- The effective annual rate is the amount by which a unit of currency will grow in a year with interest on interest included.
- An annuity is a finite set of level sequential cash flows.
- There are two types of annuities, the annuity due and the ordinary annuity. The annuity due has a first cash flow that occurs immediately; the ordinary annuity has a first cash flow that occurs one period from the present (indexed at $t = 1$).
- On a time line, we can index the present as 0 and then display equally spaced hash marks to represent a number of periods into the future. This representation allows us to index how many periods away each cash flow will be paid.
- Annuities may be handled in a similar approach as single payments if we use annuity factors rather than single-payment factors.
- The present value, PV, is the future value, FV, times the present value factor, $(1 + r)^{-N}$.
- The present value of a perpetuity is A/r, where A is the periodic payment to be received forever.
- It is possible to calculate an unknown variable, given the other relevant variables in time value of money problems.
- The cash flow additivity principle can be used to solve problems with uneven cash flows by combining single payments and annuities.

PRACTICE PROBLEMS

1. The table below gives current information on the interest rates for two two-year and two eight-year maturity investments. The table also gives the maturity, liquidity, and default risk characteristics of a new investment possibility (Investment 3). All investments promise only a single payment (a payment at maturity). Assume that premiums relating to inflation, liquidity, and default risk are constant across all time horizons.

Investment	Maturity (in Years)	Liquidity	Default Risk	Interest Rate (%)
1	2	High	Low	2.0
2	2	Low	Low	2.5
3	7	Low	Low	r_3
4	8	High	Low	4.0
5	8	Low	High	6.5

   Based on the information in the above table, address the following:
   A. Explain the difference between the interest rates on Investment 1 and Investment 2.
   B. Estimate the default risk premium.
   C. Calculate upper and lower limits for the interest rate on Investment 3, r_3.
2. A couple plans to set aside $20,000 per year in a conservative portfolio projected to earn 7 percent a year. If they make their first savings contribution one year from now, how much will they have at the end of 20 years?
3. Two years from now, a client will receive the first of three annual payments of $20,000 from a small business project. If she can earn 9 percent annually on her investments and plans to retire in six years, how much will the three business project payments be worth at the time of her retirement?
4. To cover the first year's total college tuition payments for his two children, a father will make a $75,000 payment five years from now. How much will he need to invest today to meet his first tuition goal if the investment earns 6 percent annually?
5. A client can choose between receiving 10 annual $100,000 retirement payments, starting one year from today, or receiving a lump sum today. Knowing that he can invest at a rate of 5 percent annually, he has decided to take the lump sum. What lump sum today will be equivalent to the future annual payments?
6. You are considering investing in two different instruments. The first instrument will pay nothing for three years, but then it will pay $20,000 per year for four years. The second instrument will pay $20,000 for three years and $30,000 in the fourth year. All payments are made at year-end. If your required rate of return on these investments is 8 percent annually, what should you be willing to pay for:
   A. The first instrument?
   B. The second instrument (use the formula for a four-year annuity)?

7. Suppose you plan to send your daughter to college in three years. You expect her to earn two-thirds of her tuition payment in scholarship money, so you estimate that your payments will be $10,000 a year for four years. To estimate whether you have set aside enough money, you ignore possible inflation in tuition payments and assume that you can earn 8 percent annually on your investments. How much should you set aside now to cover these payments?
8. A client plans to send a child to college for four years starting 18 years from now. Having set aside money for tuition, she decides to plan for room and board also. She estimates these costs at $20,000 per year, payable at the beginning of each year, by the time her child goes to college. If she starts next year and makes 17 payments into a savings account paying 5 percent annually, what annual payments must she make?
9. A couple plans to pay their child's college tuition for 4 years starting 18 years from now. The current annual cost of college is C$7,000, and they expect this cost to rise at an annual rate of 5 percent. In their planning, they assume that they can earn 6 percent annually. How much must they put aside each year, starting next year, if they plan to make 17 equal payments?
10. The nominal risk-free rate is *best* described as the sum of the real risk-free rate and a premium for:
    A. maturity.
    B. liquidity.
    C. expected inflation.

11. Which of the following risk premiums is most relevant in explaining the difference in yields between 30-year bonds issued by the US Treasury and 30-year bonds issued by a small private issuer?
    A. Inflation
    B. Maturity
    C. Liquidity

12. A bank quotes a stated annual interest rate of 4.00%. If that rate is equal to an effective annual rate of 4.08%, then the bank is compounding interest:
    A. daily.
    B. quarterly.
    C. semiannually.

13. The value in six years of $75,000 invested today at a stated annual interest rate of 7% compounded quarterly is *closest* to:
    A. $112,555.
    B. $113,330.
    C. $113,733.

14. A client requires £100,000 one year from now. If the stated annual rate is 2.50% compounded weekly, the deposit needed today is *closest* to:
    A. £97,500.
    B. £97,532.
    C. £97,561.

15. For a lump sum investment of ¥250,000 invested at a stated annual rate of 3% compounded daily, the number of months needed to grow the sum to ¥1,000,000 is *closest* to:
    A. 555.
    B. 563.
    C. 576.

16. Given a €1,000,000 investment for four years with a stated annual rate of 3% compounded continuously, the difference in its interest earnings compared with the same investment compounded daily is *closest* to:
    A. €1.
    B. €6.
    C. €455.

17. An investment pays €300 annually for five years, with the first payment occurring today. The present value (PV) of the investment discounted at a 4% annual rate is *closest* to:
    A. €1,336.
    B. €1,389.
    C. €1,625.

18. A perpetual preferred stock makes its first quarterly dividend payment of $2.00 in five quarters. If the required annual rate of return is 6% compounded quarterly, the stock's present value is *closest* to:
    A. $31.
    B. $126.
    C. $133.

19. A saver deposits the following amounts in an account paying a stated annual rate of 4%, compounded semiannually:

Year	End-of-Year Deposits ($)
1	4,000
2	8,000
3	7,000
4	10,000

    At the end of Year 4, the value of the account is *closest* to:
    A. $30,432
    B. $30,447
    C. $31,677

20. An investment of €500,000 today that grows to €800,000 after six years has a stated annual interest rate *closest* to:
    A. 7.5% compounded continuously.
    B. 7.7% compounded daily.
    C. 8.0% compounded semiannually.

21. A sweepstakes winner may select either a perpetuity of £2,000 a month beginning with the first payment in one month or an immediate lump sum payment of £350,000. If the annual discount rate is 6% compounded monthly, the present value of the perpetuity is:
    A. less than the lump sum.
    B. equal to the lump sum.
    C. greater than the lump sum.

22. At a 5% interest rate per year compounded annually, the present value (PV) of a 10-year ordinary annuity with annual payments of $2,000 is $15,443.47. The PV of a 10-year annuity due with the same interest rate and payments is *closest* to:
    A. $14,708.
    B. $16,216.
    C. $17,443.

23. Grandparents are funding a newborn's future university tuition costs, estimated at $50,000/year for four years, with the first payment due as a lump sum in 18 years. Assuming a 6% effective annual rate, the required deposit today is *closest* to:
    A. $60,699.
    B. $64,341.
    C. $68,201.

24. The present value (PV) of an investment with the following year-end cash flows (CF) and a 12% required annual rate of return is *closest* to:

Year	**Cash Flow (€)**
1	100,000
2	150,000
5	–10,000

    A. €201,747.
    B. €203,191.
    C. €227,573.

25. A sports car, purchased for £200,000, is financed for five years at an annual rate of 6% compounded monthly. If the first payment is due in one month, the monthly payment is *closest* to:
    A. £3,847.
    B. £3,867.
    C. £3,957.

26. Given a stated annual interest rate of 6% compounded quarterly, the level amount that, deposited quarterly, will grow to £25,000 at the end of 10 years is *closest* to:
    A. £461.
    B. £474.
    C. £836.

27. Given the following time line and a discount rate of 4% a year compounded annually, the present value (PV), as of the end of Year 5 (PV_5), of the cash flow received at the end of Year 20 is *closest* to:
    A. $22,819.
    B. $27,763.
    C. $28,873.

28. A client invests €20,000 in a four-year certificate of deposit (CD) that annually pays interest of 3.5%. The annual CD interest payments are automatically reinvested in a separate savings account at a stated annual interest rate of 2% compounded monthly. At maturity, the value of the combined asset is *closest* to:
    A. €21,670.
    B. €22,890.
    C. €22,950.

CHAPTER 2

ORGANIZING, VISUALIZING, AND DESCRIBING DATA

Pamela Peterson Drake, PhD, CFA
Jian Wu, PhD

LEARNING OUTCOMES

The candidate should be able to:

- Identify and compare data types;
- Describe how data are organized for quantitative analysis;
- Interpret frequency and related distributions;
- Interpret a contingency table;
- Describe ways that data may be visualized and evaluate uses of specific visualizations;
- Describe how to select among visualization types;
- Calculate and interpret measures of central tendency;
- Select among alternative definitions of mean to address an investment problem;
- Calculate quantiles and interpret related visualizations;
- Calculate and interpret measures of dispersion;
- Calculate and interpret target downside deviation;
- Interpret skewness;
- Interpret kurtosis;
- Interpret correlation between two variables.

1. INTRODUCTION

Data have always been a key input for securities analysis and investment management, but the acceleration in the availability and the quantity of data has also been driving the rapid evolution of the investment industry. With the rise of big data and machine learning techniques, investment practitioners are embracing an era featuring large volume, high velocity, and a wide variety of data—allowing them to explore and exploit this abundance of information for their investment strategies.

While this data-rich environment offers potentially tremendous opportunities for investors, turning data into useful information is not so straightforward. Organizing, cleaning, and analyzing data are crucial to the development of successful investment strategies; otherwise, we end up with "garbage in and garbage out" and failed investments. It is often said that 80% of an analyst's time is spent on finding, organizing, cleaning, and analyzing data, while just 20% of her/his time is taken up by model development. So, the importance of having a properly organized, cleansed, and well-analyzed dataset cannot be over-emphasized. With this essential requirement met, an appropriately executed data analysis can detect important relationships within data, uncover underlying structures, identify outliers, and extract potentially valuable insights. Utilizing both visual tools and quantitative methods, like the ones covered in this chapter, is the first step in summarizing and understanding data that will be crucial inputs to an investment strategy.

This chapter provides a foundation for understanding important concepts that are an indispensable part of the analytical tool kit needed by investment practitioners, from junior analysts to senior portfolio managers. These basic concepts pave the way for more sophisticated tools that will be developed as the quantitative methods topic unfolds and that are integral to gaining competencies in the investment management techniques and asset classes that are presented later in the CFA curriculum.

Section 2 covers core data types, including continuous and discrete numerical data, nominal and ordinal categorical data, and structured versus unstructured data. Organizing data into arrays and data tables and summarizing data in frequency distributions and contingency tables are discussed in Section 3. Section 4 introduces the important topic of data visualization using a range of charts and graphics to summarize, explore, and better understand data. Section 5 covers the key measures of central tendency, including several variants of mean that are especially useful in investments. Quantiles and their investment applications are the focus of Section 6. Key measures of dispersion are discussed in Section 7. The shape of data distributions—specifically, skewness and kurtosis—are covered in Sections 8 and 9, respectively. Section 10 provides a graphical introduction to covariance and correlation between two variables. The chapter concludes with a Summary.

2. DATA TYPES

Data can be defined as a collection of numbers, characters, words, and text—as well as images, audio, and video—in a raw or organized format to represent facts or information. To choose the appropriate statistical methods for summarizing and analyzing data and to select suitable charts for visualizing data, we need to distinguish among different data types. We will discuss data types under three different perspectives of classifications: numerical versus categorical data; cross-sectional versus time-series versus panel data; and structured versus unstructured data.

2.1. Numerical versus Categorical Data

From a statistical perspective, data can be classified into two basic groups: numerical data and categorical data.

2.1.1. Numerical Data

Numerical data are values that represent measured or counted quantities as a number and are also called **quantitative data**. Numerical (quantitative) data can be split into two types: continuous data and discrete data.

Continuous data are data that can be measured and can take on any numerical value in a specified range of values. For example, the future value of a lump-sum investment measures the amount of money to be received after a certain period of time bearing an interest rate. The future value could take on a range of values depending on the time period and interest rate. Another common example of continuous data is the price returns of a stock that measures price change over a given period in percentage terms.

Discrete data are numerical values that result from a counting process. So, practically speaking, the data are limited to a finite number of values. For example, the frequency of discrete compounding, m, counts the number of times that interest is accrued and paid out in a given year. The frequency could be monthly ($m = 12$), quarterly ($m = 4$), semi-yearly ($m = 2$), or yearly ($m = 1$).

2.1.2. Categorical Data

Categorical data (also called **qualitative data**) are values that describe a quality or characteristic of a group of observations and therefore can be used as labels to divide a dataset into groups to summarize and visualize. Usually they can take only a limited number of values that are mutually exclusive. Examples of categorical data for classifying companies include bankrupt versus not bankrupt and dividends increased versus no dividend action.

Nominal data are categorical values that are not amenable to being organized in a logical order. An example of nominal data is the classification of publicly listed stocks into 11 sectors, as shown in Exhibit 1, that are defined by the Global Industry Classification Standard (GICS). GICS, developed by Morgan Stanley Capital International (MSCI) and Standard & Poor's (S&P), is a four-tiered, hierarchical industry classification system consisting of

EXHIBIT 1 Equity Sector Classification by GICS

Sector (Text Label)	**Code** (Numerical Label)
Energy	10
Materials	15
Industrials	20
Consumer Discretionary	25
Consumer Staples	30
Health Care	35
Financials	40
Information Technology	45
Communication Services	50
Utilities	55
Real Estate	60

Source: S&P Global Market Intelligence.

11 sectors, 24 industry groups, 69 industries, and 158 sub-industries. Each sector is defined by a unique text label, as shown in the column named "Sector."

Text labels are a common format to represent nominal data, but nominal data can also be coded with numerical labels. As shown below, the column named "Code" contains a corresponding GICS code of each sector as a numerical value. However, the nominal data in numerical format do not indicate ranking, and any arithmetic operations on nominal data are not meaningful. In this example, the energy sector with the code 10 does not represent a lower or higher rank than the real estate sector with the code 60. Often, financial models, such as regression models, require input data to be numerical; so, nominal data in the input dataset must be coded numerically before applying an algorithm (that is, a process for problem solving) for performing the analysis. This would be mainly to identify the category (here, sector) in the model.

Ordinal data are categorical values that can be logically ordered or ranked. For example, the Morningstar and Standard & Poor's star ratings for investment funds are ordinal data in which one star represents a group of funds judged to have had relatively the worst performance, with two, three, four, and five stars representing groups with increasingly better performance or quality as evaluated by those firms.

Ordinal data may also involve numbers to identify categories. For example, in ranking growth-oriented investment funds based on their five-year cumulative returns, we might assign the number 1 to the top performing 10% of funds, the number 2 to next best performing 10% of funds, and so on; the number 10 represents the bottom performing 10% of funds. Despite the fact that categories represented by ordinal data can be ranked higher or lower compared to each other, they do not necessarily establish a numerical difference between each category. Importantly, such investment fund ranking tells us nothing about the difference in performance between funds ranked 1 and 2 compared with the difference in performance between funds ranked 3 and 4 or 9 and 10.

Having discussed different data types from a statistical perspective, it is important to note that at first glance, identifying data types may seem straightforward. In some situations, where categorical data are coded in numerical format, they should be distinguished from numerical data. A sound rule of thumb: Meaningful arithmetic operations can be performed on numerical data but not on categorical data.

EXAMPLE 1 Identifying Data Types (I)

Identify the data type for each of the following kinds of investment-related information:

1. *Number of coupon payments for a corporate bond.* As background, a corporate bond is a contractual obligation between an issuing corporation (i.e., borrower) and bondholders (i.e., lenders) in which the issuer agrees to pay interest—in the form of fixed coupon payments—on specified dates, typically semi-annually, over the life of the bond (i.e., to its maturity date) and to repay principal (i.e., the amount borrowed) at maturity.
2. *Cash dividends per share paid by a public company.* Note that cash dividends are a distribution paid to shareholders based on the number of shares owned.
3. *Credit ratings for corporate bond issues.* As background, credit ratings gauge the bond issuer's ability to meet the promised payments on the bond. Bond rating agencies

typically assign bond issues to discrete categories that are in descending order of credit quality (i.e., increasing probability of non-payment or default).

4. *Hedge fund classification types.* Note that hedge funds are investment vehicles that are relatively unconstrained in their use of debt, derivatives, and long and short investment strategies. Hedge fund classification types group hedge funds by the kind of investment strategy they pursue.

Solution to 1: Number of coupon payments are discrete data. For example, a newly-issued 5-year corporate bond paying interest semi-annually (quarterly) will make 10 (20) coupon payments during its life. In this case, coupon payments are limited to a finite number of values; so, they are discrete.

Solution to 2: Cash dividends per share are continuous data since they can take on any non-negative values.

Solution to 3: Credit ratings are ordinal data. A rating places a bond issue in a category, and the categories are ordered with respect to the expected probability of default. But arithmetic operations cannot be done on credit ratings, and the difference in the expected probability of default between categories of highly rated bonds, for example, is not necessarily equal to that between categories of lowly rated bonds.

Solution to 4: Hedge fund classification types are nominal data. Each type groups together hedge funds with similar investment strategies. In contrast to credit ratings for bonds, however, hedge fund classification schemes do not involve a ranking. Thus, such classification schemes are not ordinal data.

2.2. Cross-Sectional versus Time-Series versus Panel Data

Another data classification standard is based on how data are collected, and it categorizes data into three types: cross-sectional, time series, and panel.

Prior to the description of the data types, we need to explain two data-related terminologies: variable and observation. A **variable** is a characteristic or quantity that can be measured, counted, or categorized and is subject to change. A variable can also be called a field, an attribute, or a feature. For example, stock price, market capitalization, dividend and dividend yield, earnings per share (EPS), and price-to-earnings ratio (P/E) are basic data variables for the financial analysis of a public company. An **observation** is the value of a specific variable collected at a point in time or over a specified period of time. For example, last year DEF, Inc. recorded EPS of $7.50. This value represented a 15% annual increase.

Cross-sectional data are a list of the observations of a specific variable from multiple observational units at a given point in time. The observational units can be individuals, groups, companies, trading markets, regions, etc. For example, January inflation rates (i.e., the variable) for each of the euro-area countries (i.e., the observational units) in the European Union for a given year constitute cross-sectional data.

Time-series data are a sequence of observations for a single observational unit of a specific variable collected over time and at discrete and typically equally spaced intervals of time, such as daily, weekly, monthly, annually, or quarterly. For example, the daily closing prices (i.e., the variable) of a particular stock recorded for a given month constitute time-series data.

EXHIBIT 2 Earnings per Share in Euros of Three Eurozone Companies in a Given Year

Time Period	Company A	Company B	Company C
Q1	13.53	0.84	−0.34
Q2	4.36	0.96	0.08
Q3	13.16	0.79	−2.72
Q4	12.95	0.19	0.09

Panel data are a mix of time-series and cross-sectional data that are frequently used in financial analysis and modeling. Panel data consist of observations through time on one or more variables for multiple observational units. The observations in panel data are usually organized in a matrix format called a data table. Exhibit 2 is an example of panel data showing quarterly earnings per share (i.e., the variable) for three companies (i.e., the observational units) in a given year by quarter. Each column is a time series of data that represents the quarterly EPS observations from Q1 to Q4 of a specific company, and each row is cross-sectional data that represent the EPS of all three companies of a particular quarter.

2.3. Structured versus Unstructured Data

Categorizing data into structured and unstructured types is based on whether or not the data are in a highly organized form.

Structured data are highly organized in a pre-defined manner, usually with repeating patterns. The typical forms of structured data are one-dimensional arrays, such as a time series of a single variable, or two-dimensional data tables, where each column represents a variable or an observation unit and each row contains a set of values for the same columns. Structured data are relatively easy to enter, store, query, and analyze without much manual processing. Typical examples of structured company financial data are:

- Market data: data issued by stock exchanges, such as intra-day and daily closing stock prices and trading volumes.
- Fundamental data: data contained in financial statements, such as earnings per share, price to earnings ratio, dividend yield, and return on equity.
- Analytical data: data derived from analytics, such as cash flow projections or forecasted earnings growth.

Unstructured data, in contrast, are data that do not follow any conventionally organized forms. Some common types of unstructured data are text—such as financial news, posts in social media, and company filings with regulators—and also audio/video, such as managements' earnings calls and presentations to analysts.

Unstructured data are a relatively new classification driven by the rise of alternative data (i.e., data generated from unconventional sources, like electronic devices, social media, sensor networks, and satellites, but also by companies in the normal course of business) and its growing adoption in the financial industry. Unstructured data are typically alternative data as they are usually collected from unconventional sources. By indicating the source from which the data are generated, such data can be classified into three groups:

- Produced by individuals (i.e., via social media posts, web searches, etc.);

- Generated by business processes (i.e., via credit card transactions, corporate regulatory filings, etc.); and
- Generated by sensors (i.e., via satellite imagery, foot traffic by mobile devices, etc.).

Unstructured data may offer new market insights not normally contained in data from traditional sources and may provide potential sources of returns for investment processes. Unlike structured data, however, utilizing unstructured data in investment analysis is challenging. Typically, financial models are able to take only structured data as inputs; therefore, unstructured data must first be transformed into structured data that models can process.

Exhibit 3 shows an excerpt from Form 10-Q (Quarterly Report) filed by Company XYZ with the US Securities and Exchange Commission (SEC) for the fiscal quarter ended 31 March 20XX. The form is an unstructured mix of text and tables, so it cannot be directly used by computers as input to financial models. The SEC has utilized eXtensible Business Reporting Language (XBRL) to structure such data. The data extracted from the XBRL submission can be organized into five tab-delimited TXT format files that contain information about the submission, including taxonomy tags (i.e., financial statement items), dates, units of measure (uom), values (i.e., for the tag items), and more—making it readable by computer. Exhibit 4 shows an excerpt from one of the now structured data tables downloaded from the SEC's EDGAR (Electronic Data Gathering, Analysis, and Retrieval) database.

EXHIBIT 3 Excerpt from 10-Q of Company XYZ for Fiscal Quarter Ended 31 March 20XX

Company XYZ
Form 10-Q
Fiscal Quarter Ended 31 March 20XX
Table of Contents

Condensed Consolidated Statements of Operations (Unaudited) (in millions, except number of shares, which are reflected in thousands and per share amounts)	
	31 March 20XX
Net sales:	
Products	$46,565
Services	11,450
Total net sales	58,015
Cost of sales:	
Products	32,047
Services	4,147
Total cost of sales	36,194
Gross margin	21,821
Operating expenses:	
Research and development	3,948
Selling, general and administrative	4,458
Total operating expenses	8,406
Operating income	13,415
Other income/(expense), net	378
Income before provision for income taxes	13,793
Provision for income taxes	2,232
Net income	$11,561

Source: EDGAR.

EXHIBIT 4 Structured Data Extracted from Form 10-Q of Company XYZ for Fiscal Quarter Ended 31 March 20XX

adsh	tag	ddate	uom	value
0000320193-19-000066	RevenueFromContract WithCustomerExcluding AssessedTax	20XX0331	USD	$58,015,000,000
0000320193-19-000066	GrossProfit	20XX0331	USD	$21,821,000,000
0000320193-19-000066	OperatingExpenses	20XX0331	USD	$8,406,000,000
0000320193-19-000066	OperatingIncomeLoss	20XX0331	USD	$13,415,000,000
0000320193-19-000066	NetIncomeLoss	20XX0331	USD	$11,561,000,000

Source: EDGAR.

EXAMPLE 2 Identifying Data Types (II)

1. Which of the following is *most likely* to be structured data?
   A. Social media posts where consumers are commenting on what they think of a company's new product.
   B. Daily closing prices during the past month for all companies listed on Japan's Nikkei 225 stock index.
   C. Audio and video of a CFO explaining her company's latest earnings announcement to securities analysts.
2. Which of the following statements describing panel data is *most accurate*?
   A. It is a sequence of observations for a single observational unit of a specific variable collected over time at discrete and equally spaced intervals.
   B. It is a list of observations of a specific variable from multiple observational units at a given point in time.
   C. It is a mix of time-series and cross-sectional data that are frequently used in financial analysis and modeling.
3. Which of the following data series is *least likely* to be sortable by values?
   A. Daily trading volumes for stocks listed on the Shanghai Stock Exchange.
   B. EPS for a given year for technology companies included in the S&P 500 Index.
   C. Dates of first default on bond payments for a group of bankrupt European manufacturing companies.
4. Which of the following best describes a time series?
   A. Daily stock prices of the XYZ stock over a 60-month period.
   B. Returns on four-star rated Morningstar investment funds at the end of the most recent month.
   C. Stock prices for all stocks in the FTSE100 on 31 December of the most recent calendar year.

Solution to 1: B is correct as daily closing prices constitute structured data. A is incorrect as social media posts are unstructured data. C is incorrect as audio and video are unstructured data.

Solution to 2: C is correct as it most accurately describes panel data. A is incorrect as it describes time-series data. B is incorrect as it describes cross-sectional data.

Solution to 3: C is correct as dates are ordinal data that can be sorted by chronological order but not by value. A and B are incorrect as both daily trading volumes and earnings per share (EPS) are numerical data, so they can be sorted by values.

Solution to 4: A is correct since a time series is a sequence of observations of a specific variable (XYZ stock price) collected over time (60 months) and at discrete intervals of time (daily). B and C are both incorrect as they are cross-sectional data.

3. DATA SUMMARIZATION

Given the wide variety of possible formats of **raw data**, which are data available in their original form as collected, such data typically cannot be used by humans or computers to directly extract information and insights. Organizing data into a one-dimensional array or a two-dimensional array is typically the first step in data analytics and modeling. In this section, we will illustrate the construction of these typical data organization formats. We will also introduce two useful tools that can efficiently summarize one-variable and two-variable data: frequency distributions and contingency tables, respectively. Both of them can give us a quick snapshot of the data and allow us to find patterns in the data and associations between variables.

3.1. Organizing Data for Quantitative Analysis

Quantitative analysis and modeling typically require input data to be in a clean and formatted form, so raw data are usually not suitable for use directly by analysts. Depending upon the number of variables, raw data can be organized into two typical formats for quantitative analysis: one-dimensional arrays and two-dimensional rectangular arrays.

A **one-dimensional array** is the simplest format for representing a collection of data of the same data type, so it is suitable for representing a single variable. Exhibit 5 is an example of a one-dimensional array that shows the closing price for the first 10 trading days for ABC Inc. stock after the company went public. Closing prices are time-series data collected at daily intervals, so it is natural to organize them into a time-ordered sequence. The time-series format also facilitates future data updates to the existing dataset. In this case, closing prices for future trading sessions can be easily added to the end of the array with no alteration of previously formatted data.

More importantly, in contrast to compiling the data randomly in an unorganized manner, organizing such data by its time-series nature preserves valuable information beyond the basic **descriptive statistics** that summarize central tendency and spread variation in the data's distribution. For example, by simply plotting the data against time, we can learn whether the data demonstrate any increasing or decreasing trends over time or whether the time series repeats certain patterns in a systematic way over time.

A **two-dimensional rectangular array** (also called a **data table**) is one of the most popular forms for organizing data for processing by computers or for presenting data visually for consumption by humans. Similar to the structure in an Excel spreadsheet, a data table is comprised of columns and rows to hold multiple variables and multiple observations, respectively. When a data table is used to organize the data of one single observational unit (i.e., a single company), each column represents a different variable (feature or attribute) of that observational unit, and each row holds an observation for the different variables; successive rows represent the observations for successive time periods. In other words, observations of each variable are a time-series sequence that is sorted in either ascending or descending time order. Consequently, observations of different variables must be sorted and aligned to the same time scale. **Example 3** shows how to organize a raw dataset for a company collected online into a machine-readable data table.

EXHIBIT 5 One-Dimensional Array: Daily Closing Price of ABC Inc. Stock

Observation by Day	Stock Price ($)
1	57.21
2	58.26
3	58.64
4	56.19
5	54.78
6	54.26
7	56.88
8	54.74
9	52.42
10	50.14

EXAMPLE 3 Organizing a Company's Raw Data into a Data Table

Suppose you are conducting a valuation analysis of ABC Inc., which has been listed on the stock exchange for two years. The metrics to be used in your valuation include revenue, earnings per share (EPS), and dividends paid per share (DPS). You have retrieved the last two years of ABC's quarterly data from the exchange's website, which is shown in Exhibit 6. The data available online are pre-organized into a tabular format, where each column represents a fiscal year and each row represents a particular quarter with values of the three measures clustered together.

Use the data to construct a two-dimensional rectangular array (i.e., data table) with the columns representing the metrics for valuation and the observations arranged in a time-series sequence.

Solution: To construct a two-dimensional rectangular array, we first need to determine the data table structure. The columns have been specified to represent the three valuation metrics (i.e., variables): revenue, EPS and DPS. The rows should be the observations for each variable in a time ordered sequence. In this example, the data for the valuation measures will be organized in the same quarterly intervals as the raw data retrieved online, starting from Q1 Year 1 to Q4 Year 2. Then, the observations from the original table can be placed accordingly into the data table by variable name and by filing quarter. Exhibit 7 shows the raw data reorganized in the two-dimensional rectangular array (by date and associated valuation metric), which can now be used in financial analysis and is readable by a computer.

It is worth pointing out that in case of missing values while organizing data, how to handle them depends largely on why the data are missing. In this example, dividends (DPS) in the first five quarters are missing because ABC Inc. did not authorize (and pay) any dividends. So, filling the dividend column with zeros is appropriate. If revenue, EPS, and DPS of a given quarter are missing due to particular data source

EXHIBIT 6 Metrics of ABC Inc. Retrieved Online

Fiscal Quarter	Year 1 (Fiscal Year)	Year 2 (Fiscal Year)
March		
Revenue	$3,784(M)	$4,097(M)
EPS	1.37	−0.34
DPS	N/A	N/A
June		
Revenue	$4,236(M)	$5,905(M)
EPS	1.78	3.89
DPS	N/A	0.25
September		
Revenue	$4,187(M)	$4,997(M)
EPS	−3.38	−2.88
DPS	N/A	0.25
December		
Revenue	$3,889(M)	$4,389(M)
EPS	−8.66	−3.98
DPS	N/A	0.25

EXHIBIT 7 Data Table for ABC Inc.

	Revenue ($ Million)	EPS ($)	DPS ($)
Q1 Year 1	3,784	1.37	0
Q2 Year 1	4,236	1.78	0
Q3 Year 1	4,187	−3.38	0
Q4 Year 1	3,889	−8.66	0
Q1 Year 2	4,097	−0.34	0
Q2 Year 2	5,905	3.89	0.25
Q3 Year 2	4,997	−2.88	0.25
Q4 Year 2	4,389	−3.98	0.25

issues, however, these missing values cannot be simply replaced with zeros; this action would result in incorrect interpretation. Instead, the missing values might be replaced with the latest available data or with interpolated values, depending on how the data will be consumed or modeled.

3.2. Summarizing Data Using Frequency Distributions

We now discuss various tabular formats for describing data based on the count of observations. These tables are a necessary step toward building a true visualization of a dataset. Later, we shall see how bar charts, tree-maps, and heat maps, among other graphic tools, are used to visualize important properties of a dataset.

A **frequency distribution** (also called a one-way table) is a tabular display of data constructed either by counting the observations of a variable by distinct values or groups or by tallying the values of a numerical variable into a set of numerically ordered bins. It is an important tool for initially summarizing data by groups or bins for easier interpretation.

Constructing a frequency distribution of a categorical variable is relatively straightforward and can be stated in the following two basic steps:

1. Count the number of observations for each unique value of the variable.
2. Construct a table listing each unique value and the corresponding counts, and then sort the records by number of counts in descending or ascending order to facilitate the display.

Exhibit 8 shows a frequency distribution of a portfolio's stock holdings by sectors (the variables), which are defined by GICS. The portfolio contains a total of 479 stocks that have been individually classified into 11 GICS sectors (first column). The stocks are counted by sector and are summarized in the second column, absolute frequency. The **absolute frequency**, or simply the raw frequency, is the actual number of observations counted for each unique value of the variable (i.e., each sector). Often it is desirable to express the frequencies in terms of percentages, so we also show the **relative frequency** (in the third column), which is calculated as the absolute frequency of each unique value of the variable divided by the total number of observations. The relative frequency provides a normalized measure of the distribution of the data, allowing comparisons between datasets with different numbers of total observations.

EXHIBIT 8 Frequency Distribution for a Portfolio by Sector

Sector (Variable)	**Absolute Frequency**	**Relative Frequency**
Industrials	73	15.2%
Information Technology	69	14.4%
Financials	67	14.0%
Consumer Discretionary	62	12.9%
Health Care	54	11.3%
Consumer Staples	33	6.9%
Real Estate	30	6.3%
Energy	29	6.1%
Utilities	26	5.4%
Materials	26	5.4%
Communication Services	10	2.1%
Total	**479**	**100.0%**

A frequency distribution table provides a snapshot of the data, and it facilitates finding patterns. Examining the distribution of absolute frequency in Exhibit 8, we see that the largest number of stocks (73), accounting for 15.2% of the stocks in the portfolio, are held in companies in the industrials sector. The sector with the least number of stocks (10) is communication services, which represents just 2.1% of the stocks in the portfolio.

It is also easy to see that the top four sectors (i.e., industrials, information technology, financials, and consumer discretionary) have very similar relative frequencies, between 15.2% and 12.9%. Similar relative frequencies, between 6.9% and 5.4%, are also seen among several other sectors. Note that the absolute frequencies add up to the total number of stocks in the portfolio (479), and the sum of the relative frequencies should be equal to 100%.

Frequency distributions also help in the analysis of large amounts of numerical data. The procedure for summarizing numerical data is a bit more involved than that for summarizing categorical data because it requires creating non-overlapping bins (also called **intervals** or buckets) and then counting the observations falling into each bin. One procedure for constructing a frequency distribution for numerical data can be stated as follows:

1. Sort the data in ascending order.
2. Calculate the range of the data, defined as Range = Maximum value − Minimum value.
3. Decide on the number of bins (k) in the frequency distribution.
4. Determine bin width as Range/k.
5. Determine the first bin by adding the bin width to the minimum value. Then, determine the remaining bins by successively adding the bin width to the prior bin's end point and stopping after reaching a bin that includes the maximum value.
6. Determine the number of observations falling into each bin by counting the number of observations whose values are equal to or exceed the bin minimum value yet are less than the bin's maximum value. The exception is in the last bin, where the maximum value is equal to the last bin's maximum, and therefore, the observation with the maximum value is included in this bin's count.
7. Construct a table of the bins listed from smallest to largest that shows the number of observations falling into each bin.

In Step 4, when rounding the bin width, round up (rather than down) to ensure that the final bin includes the maximum value of the data.

These seven steps are basic guidelines for constructing frequency distributions. In practice, however, we may want to refine the above basic procedure. For example, we may want the bins to begin and end with whole numbers for ease of interpretation. Another practical refinement that promotes interpretation is to start the first bin at the nearest whole number below the minimum value.

As this procedure implies, a frequency distribution groups data into a set of bins, where each bin is defined by a unique set of values (i.e., beginning and ending points). Each observation falls into only one bin, and the total number of bins covers all the values represented in the data. The frequency distribution is the list of the bins together with the corresponding measures of frequency.

To illustrate the basic procedure, suppose we have 12 observations sorted in ascending order (*Step 1*):

−4.57, −4.04, −1.64, 0.28, 1.34, 2.35, 2.38, 4.28, 4.42, 4.68, 7.16, and 11.43.

The minimum observation is −4.57, and the maximum observation is +11.43. So, the range is +11.43 − (−4.57) = 16 (*Step 2*).

EXHIBIT 9 Determining Endpoints of the Bins

−4.57	+	4.0	=	−0.57
−0.57	+	4.0	=	3.43
3.43	+	4.0	=	7.43
7.40	+	4.0	=	11.43

If we set $k = 4$ (*Step 3*), then the bin width is $16/4 = 4$ (*Step 4*).

Exhibit 9 shows the repeated addition of the bin width of 4 to determine the endpoint for each of the bins (*Step 5*).

Thus, the bins are [−4.57 to −0.57), [−0.57 to 3.43), [3.43 to 7.43), and [7.43 to 11.43], where the notation [−4.57 to −0.57) indicates $-4.57 \leq$ observation < -0.57. The parentheses indicate that the endpoints are not included in the bins, and the square brackets indicate that the beginning points and the last endpoint are included in the bin. Exhibit 10 summarizes Steps 5 through 7.

Note that the bins do not overlap, so each observation can be placed uniquely into one bin, and the last bin includes the maximum value.

We turn to these issues in discussing the construction of frequency distributions for daily returns of the fictitious Euro-Asia-Africa (EAA) Equity Index. The dataset of daily returns of the EAA Equity Index spans a five-year period and consists of 1,258 observations with a minimum value of −4.1% and a maximum value of 5.0%. Thus, the range of the data is 5% − (−4.1%) = 9.1%, approximately. (The mean daily return—mean as a measure of central tendency will be discussed shortly—is 0.04%.)

The decision on the number of bins (k) into which we should group the observations often involves inspecting the data and exercising judgment. How much detail should we include? If we use too few bins, we will summarize too much and may lose pertinent characteristics. Conversely, if we use too many bins, we may not summarize enough and may introduce unnecessary noise.

We can establish an appropriate value for k by evaluating the usefulness of the resulting bin width. A large number of empty bins may indicate that we are attempting to over-organize the data to present too much detail. Starting with a relatively small bin width, we can see whether or not the bins are mostly empty and whether or not the value of k associated with that bin width is too large. If the bins are mostly empty, implying that k is too large, we can consider increasingly larger bins (i.e., smaller values of k) until we have a frequency distribution that effectively summarizes the distribution.

EXHIBIT 10 Frequency Distribution

Bin				**Absolute Frequency**
A	−4.57	≤ observation <	−0.57	3
B	−0.57	≤ observation <	3.43	4
C	3.43	≤ observation <	7.43	4
D	7.43	≤ observation ≤	11.43	1

Suppose that for ease of interpretation we want to use a bin width stated in whole rather than fractional percentages. In the case of the daily EAA Equity Index returns, a 1 percent bin width would be associated with $9.1/1 = 9.1$ bins, which can be rounded up to $k = 10$ bins. That number of bins will cover a range of $1\% \times 10 = 10\%$. By constructing the frequency distribution in this manner, we will also have bins that end and begin at a value of 0 percent, thereby allowing us to count the negative and positive returns in the data. Without too much work, we have found an effective way to summarize the data.

Exhibit 11 shows the frequency distribution for the daily returns of the EAA Equity Index using return bins of 1 percent, where the first bin includes returns from -5.0 percent to -4.0 percent (exclusive, meaning < -4 percent) and the last bin includes daily returns from 4.0 percent to 5.0 percent (inclusive, meaning ≤ 5 percent). Note that to facilitate interpretation, the first bin starts at the nearest whole number below the minimum value (so, at -5.0 percent).

Exhibit 11 includes two other useful ways to present the data (which can be computed in a straightforward manner once we have established the absolute and relative frequency distributions): the cumulative absolute frequency and the cumulative relative frequency. The **cumulative absolute frequency** cumulates (meaning, adds up) the absolute frequencies as we move from the first bin to the last bin. Similarly, the **cumulative relative frequency** is a sequence of partial sums of the relative frequencies. For the last bin, the cumulative absolute frequency will equal the number observations in the dataset (1,258), and the cumulative relative frequency will equal 100 percent.

As Exhibit 11 shows, the absolute frequencies vary widely, ranging from 1 to 555. The bin encompassing returns between 0 percent and 1 percent has the most observations (555), and the corresponding relative frequency tells us these observations account for 44.12 percent of the total number of observations. The frequency distribution gives us a sense of not only where most of the observations lie but also whether the distribution is evenly spread. It is easy to see that the vast majority of observations (37.36 percent + 44.12 percent = 81.48 percent) lie in the middle two bins spanning -1 percent to 1 percent. We can also see that not many observations are greater than 3 percent or less than -4 percent. Moreover, as there are bins

EXHIBIT 11 Frequency Distribution for Daily Returns of EAA Equity Index

Return Bin (%)	Absolute Frequency	Relative Frequency (%)	Cumulative Absolute Frequency	Cumulative Relative Frequency (%)
−5.0 to −4.0	1	0.08	1	0.08
−4.0 to −3.0	7	0.56	8	0.64
−3.0 to −2.0	23	1.83	31	2.46
−2.0 to −1.0	77	6.12	108	8.59
−1.0 to 0.0	470	37.36	578	45.95
0.0 to 1.0	555	44.12	1,133	90.06
1.0 to 2.0	110	8.74	1,243	98.81
2.0 to 3.0	13	1.03	1,256	99.84
3.0 to 4.0	1	0.08	1,257	99.92
4.0 to 5.0	1	0.08	1,258	100.00

with 0 percent as ending or beginning points, we are able to count positive and negative returns in the data. Looking at the cumulative relative frequency in the last column, we see that the bin of −1 percent to 0 percent shows a cumulative relative frequency of 45.95 percent. This indicates that 45.95 percent of the observations lie below the daily return of 0 percent and that 54.05 percent of the observations are positive daily returns.

It is worth noting that other than being summarized in tables, frequency distributions also can be effectively represented in visuals, which will be discussed shortly in the section on data visualization.

EXAMPLE 4 Constructing a Frequency Distribution of Country Index Returns

Suppose we have the annual equity index returns of a given year for 18 different countries, as shown in Exhibit 12, and we are asked to summarize the data.

Construct a frequency distribution table from these data and state some key findings from the summarized data.

Solution: The first step in constructing a frequency distribution table is to sort the return data in ascending order:
The second step is to calculate the range of the data, which is 9.9% − 5.5% = 4.4%.

EXHIBIT 12 Annual Equity Index Returns for 18 Countries

Market	**Index Return (%)**
Country A	7.7
Country B	8.5
Country C	9.1
Country D	5.5
Country E	7.1
Country F	9.9
Country G	6.2
Country H	6.8
Country I	7.5
Country J	8.9
Country K	7.4
Country L	8.6
Country M	9.6
Country N	7.7
Country O	6.8
Country P	6.1
Country Q	8.8
Country R	7.9

Market	Index Return (%)
Country D	5.5
Country P	6.1
Country G	6.2
Country H	6.8
Country O	6.8
Country E	7.1
Country K	7.4
Country I	7.5
Country A	7.7
Country N	7.7
Country R	7.9
Country B	8.5
Country L	8.6
Country Q	8.8
Country J	8.9
Country C	9.1
Country M	9.6
Country F	9.9

The third step is to decide on the number of bins. Here, we will use $k = 5$.
The fourth step is to determine the bin width. Here, it is 4.4%/5 = 0.88%, which we will round up to 1.0%.
The fifth step is to determine the bins, which are as follows:

5.0% + 1.0% = 6.0%
6.0% + 1.0% = 7.0%
7.0% + 1.0% = 8.0%
8.0% + 1.0% = 9.0%
9.0% + 1.0% = 10.0%

For ease of interpretation, the first bin is set to begin with the nearest whole number (5.0 percent) below the minimum value (5.5 percent) of the data series.
The sixth step requires counting the return observations falling into each bin, and the seventh (last) step is use these results to construct the final frequency distribution table.
Exhibit 13 presents the frequency distribution table, which summarizes the data in Exhibit 12 into five bins spanning 5 to 10 percent. Note that with 18 countries, the relative frequency for one observation is calculated as 1/18 = 5.56 percent.

As Exhibit 13 shows, there is substantial variation in these equity index returns. One-third of the observations fall in the 7.0 to 8.0% bin, making it the bin with the most observations. Both the 6.0 to 7.0% bin and the 8.0 to 9.0% bin hold four

EXHIBIT 13 Frequency Distribution of Equity Index Returns

Return Bin (%)	Absolute Frequency	Relative Frequency (%)	Cumulative Absolute Frequency	Cumulative Relative Frequency (%)
5.0 to 6.0	1	5.56	1	5.56
6.0 to 7.0	4	22.22	5	27.78
7.0 to 8.0	6	33.33	11	61.11
8.0 to 9.0	4	22.22	15	83.33
9.0 to 10.0	3	16.67	18	100.00

observations each, accounting for 22.22 percent of the total number of the observations, respectively. The two remaining bins have fewer observations, one or three observations, respectively.

3.3. Summarizing Data Using a Contingency Table

We have shown that the frequency distribution table is a powerful tool to summarize data for one variable. How can we summarize data for two variables simultaneously? A contingency table provides a solution to this question.

A **contingency table** is a tabular format that displays the frequency distributions of two or more categorical variables simultaneously and is used for finding patterns between the variables. A contingency table for two categorical variables is also known as a two-way table. Contingency tables are constructed by listing all the levels (i.e., categories) of one variable as rows and all the levels of the other variable as columns in the table. A contingency table having R levels of one variable in rows and C levels of the other variable in columns is referred to as an $R \times C$ table. Note that each variable in a contingency table must have a finite number of levels, which can be either ordered (ordinal data) or unordered (nominal data). Importantly, the data displayed in the cells of the contingency table can be either a frequency (count) or a relative frequency (percentage) based on either overall total, row totals, or column totals.

Exhibit 14 presents a 5×3 contingency table that summarizes the number of stocks (i.e., frequency) in a particular portfolio of 1,000 stocks by two variables, sector and company market capitalization. Sector has five levels, with each one being a GICS-defined sector. Market capitalization (commonly referred to as market cap) is defined for a company as the number of shares outstanding times the price per share. The stocks in this portfolio are categorized by three levels of market capitalization: large cap, more than \$10 billion; mid cap, \$10 billion to \$2 billion; and small cap, less than \$2 billion.

The entries in the cells of the contingency table show the number of stocks of each sector with a given level of market cap. For example, there are 275 small-cap health care stocks, making it the portfolio's largest subgroup in terms of frequency. These data are also called **joint frequencies** because you are joining one variable from the row (i.e., sector) and the other variable from the column (i.e., market cap) to count observations. The joint frequencies are then added across rows and across columns, and these corresponding sums

EXHIBIT 14 Portfolio Frequencies by Sector and Market Capitalization

	Market Capitalization Variable (3 Levels)			
Sector Variable (5 Levels)	**Small**	**Mid**	**Large**	**Total**
Communication Services	55	35	20	110
Consumer Staples	50	30	30	110
Energy	175	95	20	290
Health Care	275	105	55	435
Utilities	20	25	10	55
Total	**575**	**290**	**135**	**1,000**

are called **marginal frequencies**. For example, the marginal frequency of health care stocks in the portfolio is the sum of the joint frequencies across all three levels of market cap, so 435 (= 275 + 105 + 55). Similarly, adding the joint frequencies of small-cap stocks across all five sectors gives the marginal frequency of small-cap stocks of 575 (= 55 + 50 + 175 + 275 + 20).

Clearly, health care stocks and small-cap stocks have the largest marginal frequencies among sector and market cap, respectively, in this portfolio. Note the marginal frequencies represent the frequency distribution for each variable. Finally, the marginal frequencies for each variable must sum to the total number of stocks (overall total) in the portfolio—here, 1,000 (shown in the lower right cell).

Similar to the one-way frequency distribution table, we can express frequency in percentage terms as relative frequency by using one of three options. We can divide the joint frequencies by: a) the total count; b) the marginal frequency on a row; or c) the marginal frequency on a column.

Exhibit 15 shows the contingency table using relative frequencies based on total count. It is readily apparent that small-cap health care and energy stocks comprise the largest portions of the total portfolio, at 27.5 percent (= 275/1,000) and 17.5 percent (= 175/1,000),

EXHIBIT 15 Relative Frequencies as Percentage of Total

	Market Capitalization Variable (3 Levels)			
Sector Variable (5 Levels)	**Small**	**Mid**	**Large**	**Total**
Communication Services	5.5%	3.5%	2.0%	11.0%
Consumer Staples	5.0%	3.0%	3.0%	11.0%
Energy	17.5%	9.5%	2.0%	29.0%
Health Care	27.5%	10.5%	5.5%	43.5%
Utilities	2.0%	2.5%	1.0%	5.5%
Total	**57.5%**	**29.0%**	**13.5%**	**100%**

EXHIBIT 16 Relative Frequencies: Sector as Percentage of Market Cap

	Market Capitalization Variable (3 Levels)			
Sector Variable (5 Levels)	**Small**	**Mid**	**Large**	**Total**
Communication Services	9.6%	12.1%	14.8%	11.0%
Consumer Staples	8.7%	10.3%	22.2%	11.0%
Energy	30.4%	32.8%	14.8%	29.0%
Health Care	47.8%	36.2%	40.7%	43.5%
Utilities	3.5%	8.6%	7.4%	5.5%
Total	100.0%	100.0%	100.0%	**100.0%**

respectively, followed by mid-cap health care and energy stocks, at 10.5 percent and 9.5 percent, respectively. Together, these two sectors make up nearly three-quarters of the portfolio (43.5% + 29.0% = 72.5%).

Exhibit 16 shows relative frequencies based on marginal frequencies of market cap (i.e., columns). From this perspective, it is clear that the health care and energy sectors dominate the other sectors at each level of market capitalization: 78.3 percent (= 275/575 + 175/575), 69.0 percent (= 105/290 + 95/290), and 55.6 percent (= 55/135 + 20/135), for small, mid, and large caps, respectively. Note that there may be a small rounding error difference between these results and the numbers shown in Exhibit 15.

In conclusion, the findings from these contingency tables using frequencies and relative frequencies indicate that in terms of the number of stocks, the portfolio can be generally described as a small- to mid-cap–oriented health care and energy sector portfolio that also includes stocks of several other defensive sectors.

As an analytical tool, contingency tables can be used in different applications. One application is for evaluating the performance of a classification model (in this case, the contingency table is called a **confusion matrix**). Suppose we have a model for classifying companies into two groups: those that default on their bond payments and those that do not default. The confusion matrix for displaying the model's results will be a 2 × 2 table showing the frequency of actual defaults versus the model's predicted frequency of defaults. Exhibit 17 shows such a confusion matrix for a sample of 2,000 non-investment-grade bonds. Using company characteristics and other inputs, the model correctly predicts 300 cases of bond defaults and 1,650 cases of no defaults.

EXHIBIT 17 Confusion Matrix for Bond Default Prediction Model

Predicted	**Actual Default**		
Default	**Yes**	**No**	**Total**
Yes	300	40	**340**
No	10	1,650	**1,660**
Total	**310**	**1,690**	**2,000**

We can also observe that this classification model incorrectly predicts default in 40 cases where no default actually occurred and also incorrectly predicts no default in 10 cases where default actually did occur. Later in the text you will learn how to construct a confusion matrix, how to calculate related model performance metrics, and how to use them to evaluate and tune a classification model.

Another application of contingency tables is to investigate potential association between two categorical variables. For example, revisiting Exhibit 14, one may ask whether the distribution of stocks by sectors is independent of the levels of market capitalization? Given the dominance of small-cap and mid-cap health care and energy stocks, the answer is likely, no.

One way to test for a potential association between categorical variables is to perform a **chi-square test of independence**. Essentially, the procedure involves using the marginal frequencies in the contingency table to construct a table with expected values of the observations. The actual values and expected values are used to derive the chi-square test statistic. This test statistic is then compared to a value from the chi-square distribution for a given level of significance. If the test statistic is greater than the chi-square distribution value, then there is evidence to reject the claim of independence, implying a significant association exists between the categorical variables. The following example describes how a contingency table is used to set up this test of independence.

EXAMPLE 5 Contingency Tables and Association between Two Categorical Variables

Suppose we randomly pick 315 investment funds and classify them two ways: by fund style, either a growth fund or a value fund; and by risk level, either low risk or high risk. Growth funds primarily invest in stocks whose earnings are expected to grow at a faster rate than earnings for the broad stock market. Value funds primarily invest in stocks that appear to be undervalued relative to their fundamental values. Risk here refers to volatility in the return of a given investment fund, so low (high) volatility implies low (high) risk. The data are summarized in a 2×2 contingency table shown in Exhibit 18.

1. Calculate the number of growth funds and number of value funds out of the total funds.
2. Calculate the number of low-risk and high-risk funds out of the total funds.
3. Describe how the contingency table is used to set up a test for independence between fund style and risk level.

Solution to 1: The task is to calculate the marginal frequencies by fund style, which is done by adding joint frequencies across the rows. Therefore, the marginal frequency for growth is $73 + 26 = 99$, and the marginal frequency for value is $183 + 33 = 216$.

EXHIBIT 18 Contingency Table by Investment Fund Style and Risk Level

	Low Risk	**High Risk**
Growth	73	26
Value	183	33

Solution to 2: The task is to calculate the marginal frequencies by fund risk, which is done by adding joint frequencies down the columns. Therefore, the marginal frequency for low risk is 73 + 183 = 256, and the marginal frequency for high risk is 26 + 33 = 59.

Solution to 3: Based on the procedure mentioned for conducting a chi-square test of independence, we would perform the following three steps.

Step 1: Add the marginal frequencies and overall total to the contingency table. We have also included the relative frequency table for observed values.

EXHIBIT 19A Observed Marginal Frequencies and Relative Frequencies

	Observed Values				**Observed Values**		
	Low Risk	**High Risk**			**Low Risk**	**High Risk**	
Growth	73	26	**99**	Growth	74%	26%	**100%**
Value	183	33	**216**	Value	85%	15%	**100%**
	256	**59**	**315**				

Step 2: Use the marginal frequencies in the contingency table to construct a table with expected values of the observations. To determine expected values for each cell, multiply the respective row total by the respective column total, then divide by the overall total. So, for $cell_{i,j}$ (in *ith* row and *jth* column):

$$\text{Expected Value}_{i,j} = (\text{Total Row } i \times \text{Total Column } j)/\text{Overall Total} \tag{1}$$

For example,

Expected value for Growth/Low Risk is: $(99 \times 256)/315 = 80.46$; and

Expected value for Value/High Risk is: $(216 \times 59)/315 = 40.46$.

The table of expected values (and accompanying relative frequency table) are:

EXHIBIT 19B Expected Marginal Frequencies and Relative Frequencies

	Observed Values				**Observed Values**		
	Low Risk	**High Risk**			**Low Risk**	**High Risk**	
Growth	80.457	18.543	**99**	Growth	81%	19%	**100%**
Value	175.543	40.457	**216**	Value	81%	19%	**100%**
	256	**59**	**315**				

Step 3: Use the actual values and the expected values of observation counts to derive the chi-square test statistic, which is then compared to a value from the chi-square distribution for a given level of significance. If the test statistic is greater than the chi-square distribution value, then there is evidence of a significant association between the categorical variables.

4. DATA VISUALIZATION

Visualization is the presentation of data in a pictorial or graphical format for the purpose of increasing understanding and for gaining insights into the data. As has been said, "a picture is worth a thousand words." In this section, we discuss a variety of charts that are useful for understanding distributions, making comparisons, and exploring potential relationships among data. Specifically, we will cover visualizing frequency distributions of numerical and categorical data by using plots that represent multi-dimensional data for discovering relationships and by interpreting visuals that display unstructured data.

4.1. Histogram and Frequency Polygon

A **histogram** is a chart that presents the distribution of numerical data by using the height of a bar or column to represent the absolute frequency of each bin or interval in the distribution.

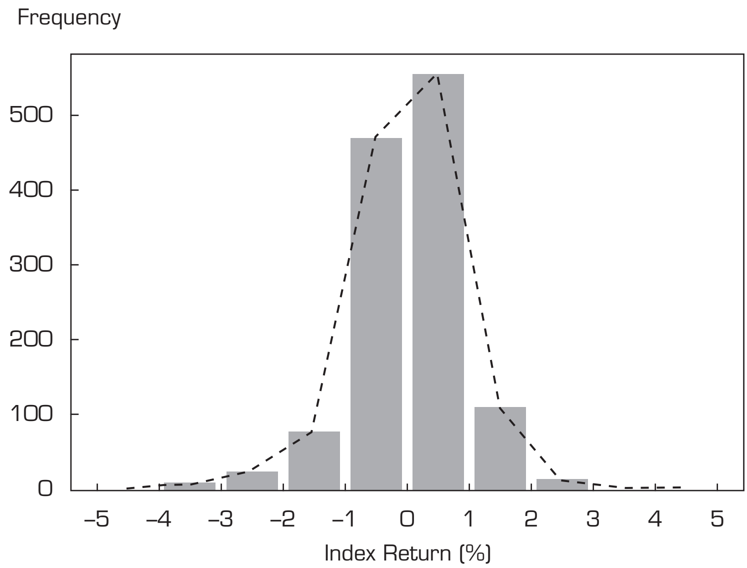
To construct a histogram from a continuous variable, we first need to split the data into bins and summarize the data into a frequency distribution table, such as the one we constructed in Exhibit 11. In a histogram, the *y*-axis generally represents the absolute frequency or the relative frequency in percentage terms, while the *x*-axis usually represents the bins of the variable. Using the frequency distribution table in Exhibit 11, we plot the histogram of daily returns of the EAA Equity Index, as shown in Exhibit 20. The bars are of equal width, representing the bin width of 1 percent for each return interval. The bars are usually drawn with no spaces in between, but small gaps can also be added between adjacent bars to increase readability, as in this exhibit. In this case, the height of each bar represents the absolute frequency for each return bin. A quick glance can tell us that the return bin 0 to 1 percent (exclusive) has the highest frequency, with more than 500 observations (555, to be exact), and it is represented by the tallest bar in the histogram.

An advantage of the histogram is that it can effectively present a large amount of numerical data that has been grouped into a frequency distribution and can allow a quick inspection of the shape, center, and spread of the distribution to better understand it. For example, in Exhibit 20, despite the histogram of daily EAA Equity Index returns appearing bell-shaped and roughly symmetrical, most bars to the right side of the origin (i.e., zero) are taller than those on the left side, indicating that more observations lie in the bins in positive territory. Remember that in the earlier discussion of this return distribution, it was noted that 54.1 percent of the observations are positive daily returns.

As mentioned, histograms can also be created with relative frequencies—the choice of using absolute versus relative frequency depends on the question being answered. An absolute frequency histogram best answers the question of how many items are in each bin, while a relative frequency histogram gives the proportion or percentage of the total observations in each bin.

Another graphical tool for displaying frequency distributions is the frequency polygon. To construct a **frequency polygon**, we plot the midpoint of each return bin on the *x*-axis and the absolute frequency for that bin on the *y*-axis. We then connect neighboring points with a straight line. Exhibit 20 shows the frequency polygon that overlays the histogram. In the graph, for example, the return interval 1 to 2 percent (exclusive) has a frequency of 110, so we plot the return-interval midpoint of 0.5 percent (which is 1.50 percent on the *x*-axis) and a frequency of 110 (on the *y*-axis). Importantly, the frequency polygon can quickly convey a visual understanding of the distribution since it displays frequency as an area under the curve.

EXHIBIT 20 Histogram Overlaid with Frequency Polygon for Daily Returns of EAA Equity Index



Another form for visualizing frequency distributions is the **cumulative frequency distribution chart**. Such a chart can plot either the cumulative absolute frequency or the cumulative relative frequency on the *y*-axis against the upper limit of the interval. The cumulative frequency distribution chart allows us to see the number or the percentage of the observations that lie below a certain value. To construct the cumulative frequency distribution, we graph the returns in the fourth (i.e., Cumulative Absolute Frequency) or fifth (i.e., Cumulative Relative Frequency) column of Exhibit 11 against the upper limit of each return interval.

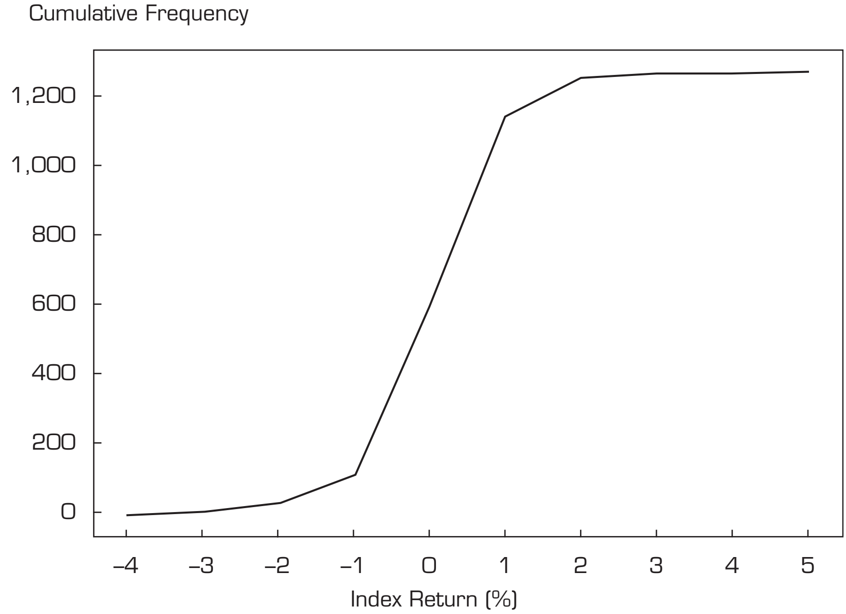
Exhibit 21 presents the graph of the cumulative absolute frequency distribution for the daily returns on the EAA Equity Index. Notice that the cumulative distribution tends to flatten out when returns are extremely negative or extremely positive because the frequencies in these bins are quite small. The steep slope in the middle of Exhibit 21 reflects the fact that most of the observations—[(470 + 555)/1,258], or 81.5 percent—lie in the neighborhood of −1.0 to 1.0 percent.

4.2. Bar Chart

As we have demonstrated, the histogram is an efficient graphical tool to present the frequency distribution of numerical data. The frequency distribution of categorical data can be plotted in a similar type of graph called a **bar chart**. In a bar chart, each bar represents a distinct category, with the bar's height proportional to the frequency of the corresponding category.

Similar to plotting a histogram, the construction of a bar chart with one categorical variable first requires a frequency distribution table summarized from the variable. Note that the bars can be plotted vertically or horizontally. In a vertical bar chart, the *y*-axis still represents the absolute frequency or the relative frequency. Different from the histogram,

EXHIBIT 21 Cumulative Absolute Frequency Distribution of Daily Returns of EAA Equity Index



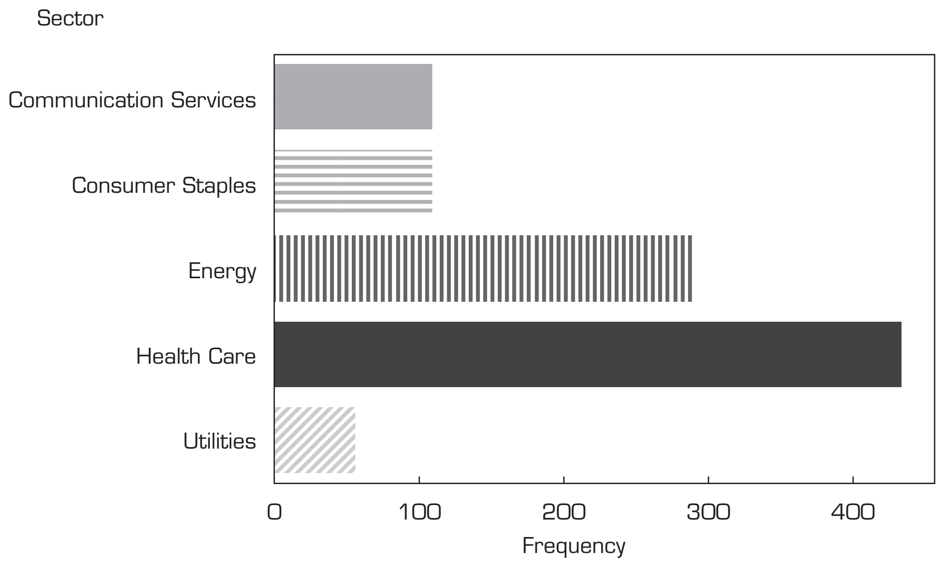
however, is that the *x*-axis in a bar chart represents the mutually exclusive categories to be *compared* rather than bins that group numerical data.

For example, using the marginal frequencies for the five GICS sectors shown in the last column in Exhibit 14, we plot a horizontal bar chart in Exhibit 22 to show the frequency of stocks by sector in the portfolio. The bars are of equal width to represent each sector, and sufficient space should be between adjacent bars to separate them from each other. Because this is a horizontal bar chart—in this case, the *x*-axis shows the absolute frequency and the *y*-axis represents the sectors—the length of each bar represents the absolute frequency of each sector. Since sectors are nominal data with no logical ordering, the bars representing sectors may be arranged in any order. However, in the particular case where the categories in a bar chart are ordered by frequency in descending order and the chart includes a line displaying cumulative relative frequency, then it is called a Pareto Chart. The chart is often used to highlight dominant categories or the most important groups.

Bar charts provide a snapshot to show the comparison between categories of data. As shown in Exhibit 22, the sector in which the portfolio holds most stocks is the health care sector, with 435 stocks, followed by the energy sector, with 290 stocks. The sector in which the portfolio has the least number of stocks is utilities, with 55 stocks. To compare categories more accurately, in some cases we may add the frequency count to the right end of each bar (or the top end of each bar in the case of a vertical bar chart).

The bar chart shown in Exhibit 22 can present the frequency distribution of only one categorical variable. In the case of two categorical variables, we need an enhanced version of the bar chart, called a **grouped bar chart** (also known as a **clustered bar chart**), to show joint frequencies. Using the joint frequencies by sector and by level of market capitalization given

EXHIBIT 22 Frequency by Sector for Stocks in a Portfolio



in Exhibit 14, for example, we show how a grouped bar chart is constructed in Exhibit 23. While the *y*-axis still represents the same categorical variable (the distinct GICS sectors as in Exhibit 22), in Exhibit 23 three bars are clustered side-by-side within the same sector to represent the three respective levels of market capitalization. The bars within each cluster should be colored or patterned differently to distinguish between them, but the color or pattern schemes for the sub-groups must be identical across the sector clusters, as shown by the legend at the upper right of Exhibit 23. Additionally, the bars in each sector cluster must always be placed in the same order throughout the chart. It is easy to see that the small-cap heath care stocks are the sub-group with the highest frequency (275), and we can also see that small-cap stocks are the largest sub-group within each sector—except for utilities, where mid cap is the largest.

An alternative form for presenting the joint frequency distribution of two categorical variables is a **stacked bar chart**. In the vertical version of a stacked bar chart, the bars representing the sub-groups are placed on top of each other to form a single bar. Each subsection of the bar is shown in a different color or pattern to represent the contribution of each sub-group, and the overall height of the stacked bar represents the marginal frequency for the category. Exhibit 23 can be replotted in a stacked bar chart, as shown in Exhibit 24.

We have shown that the frequency distribution of categorical data can be clearly and efficiently presented by using a bar chart. However, it is worth noting that applications of bar charts may be extended to more general cases when categorical data are associated with numerical data. For example, suppose we want to show a company's quarterly profits over the past one year. In this case, we can plot a vertical bar chart where each bar represents one of the four quarters in a time order and its height indicates the value of profits for that quarter.

EXHIBIT 23 Frequency by Sector and Level of Market Capitalization for Stocks in a Portfolio

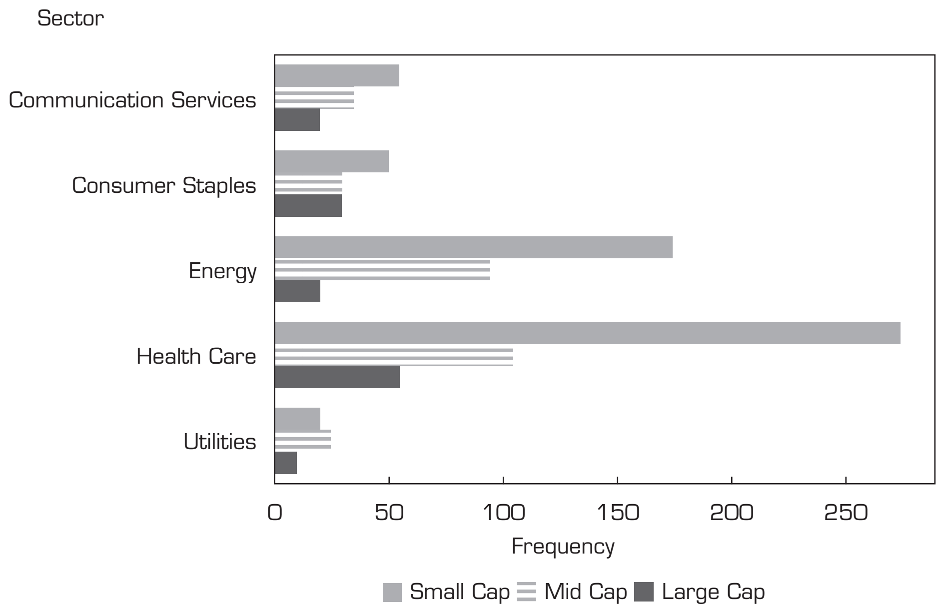


EXHIBIT 24 Frequency by Sector and Level of Market Capitalization in a Stacked Bar Chart

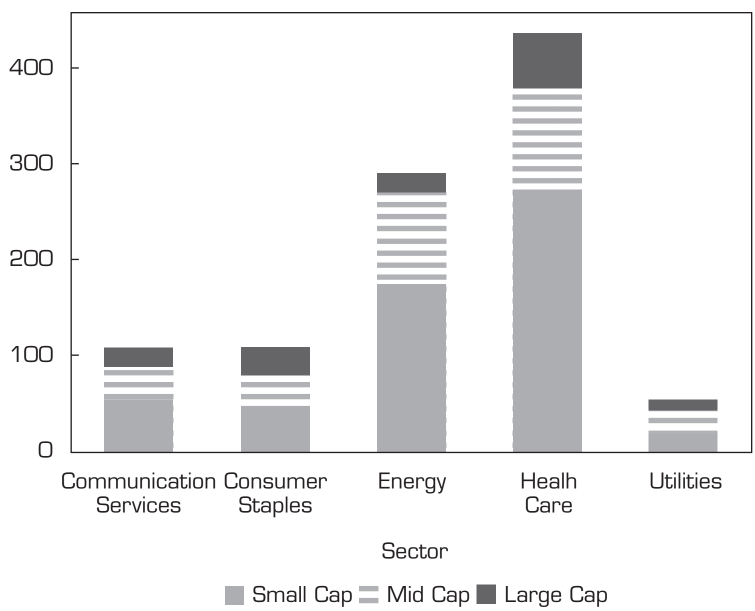
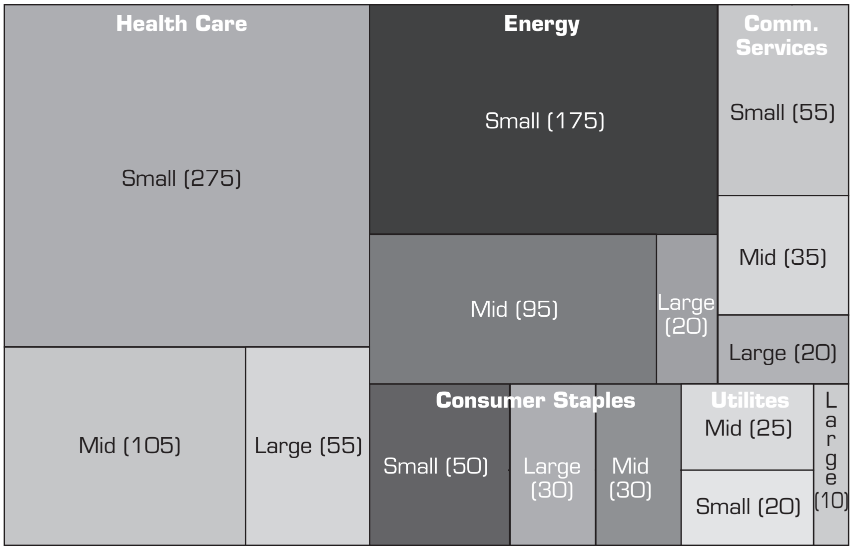


EXHIBIT 25 Tree-Map for Frequency Distribution by Sector in a Portfolio



4.3. Tree-Map

In addition to bar charts and grouped bar charts, another graphical tool for displaying categorical data is a **tree-map**. It consists of a set of shaded rectangles to represent distinct groups, and the area of each rectangle is proportional to the value of the corresponding group. For example, referring back to the marginal frequencies by GICS sector in Exhibit 14, we plot a tree-map in Exhibit 25 to represent the frequency distribution by sector for stocks in the portfolio. The tree-map clearly shows that health care is the sector with the largest number of stocks in the portfolio, which is represented by the rectangle with the largest area.

Note that this example also depicts one more categorical variable (i.e., level of market capitalization). The tree-map can represent data with additional dimensions by displaying a set of nested rectangles. To show the joint frequencies of sub-groups by sector and level of market capitalization, as given in Exhibit 14, we can split each existing rectangle for sector into three sub-rectangles to represent small-cap, mid-cap, and large-cap stocks, respectively. In this case, the area of each nested rectangle would be proportional to the number of stocks in each market capitalization sub-group. The exhibit clearly shows that small-cap health care is the sub-group with the largest number of stocks. It is worth noting a caveat for using tree-maps: Tree-maps become difficult to read if the hierarchy involves more than three levels.

4.4. Word Cloud

So far, we have shown how to visualize the frequency distribution of numerical data or categorical data. However, can we find a chart to depict the frequency of unstructured data—particularly, textual data? A **word cloud** (also known as **tag cloud**) is a visual device for representing textual data. A word cloud consists of words extracted from a source of textual data, with the size of each distinct word being proportional to the frequency with which it appears in the given text. Note that common words (e.g., "a," "it," "the") are generally stripped out to focus on key words that convey the most meaningful information. This format allows us to quickly perceive the most frequent terms among the given text to provide information about the nature of the text, including topic and whether or not the text conveys positive or negative news. Moreover, words conveying different sentiment may be displayed in different colors. For example, "profit" typically indicates positive sentiment so might be displayed in green, while "loss" typically indicates negative sentiment and may be shown in red.

Exhibit 26 is an excerpt from the Management's Discussion and Analysis (MDA) section of the 10-Q filing for QXR Inc. for the quarter ended 31 March 20XX. Taking this text, we can create a word cloud, as shown in Exhibit 27. A quick glance at the word cloud tells us that the following words stand out (i.e., they were used most frequently in the MDA text): "billion," "revenue," "year," "income," "growth," and "financial." Note that specific words, such as "income" and "growth," typically convey positive sentiment, as contrasted with such words as "loss" and "decline," which typically convey negative sentiment. In conclusion, word clouds are a useful tool for visualizing textual data that can facilitate understanding the topic of the text as well as the sentiment it may convey.

EXHIBIT 26 Excerpt of MDA Section in Form 10-Q of QXR Inc. for Quarter Ended 31 March 20XX

MANAGEMENT'S DISCUSSION AND ANALYSIS OF FINANCIAL CONDITION AND RESULTS OF OPERATIONS

Please read the following discussion and analysis of our financial condition and results of operations together with our consolidated financial statements and related notes included under Part I, Item 1 of this Quarterly Report on Form 10-Q.

EXECUTIVE OVERVIEW OF RESULTS

Below are our key financial results for the three months ended March 31, 20XX (consolidated unless otherwise noted):

- Revenues of $36.3 billion and revenue growth of 17 percent year over year, constant currency revenue growth of 19 percent year over year.
- Major segment revenues of $36.2 billion with revenue growth of 17 percent year over year and other segments' revenues of $170 million with revenue growth of 13 percent year over year.
- Revenues from the United States, EMEA, APAC, and Other Americas were $16.5 billion, $11.8 billion, $6.1 billion, and $1.9 billion, respectively.
- Cost of revenues was $16.0 billion, consisting of TAC of $6.9 billion and other cost of revenues of $9.2 billion. Our TAC as a percentage of advertising revenues were 22 percent.
- Operating expenses (excluding cost of revenues) were $13.7 billion, including the EC AFS fine of $1.7 billion.
- Income from operations was $6.6 billion.
- Other income (expense), net, was $1.5 billion.
- Effective tax rate was 18 percent
- Net income was $6.7 billion with diluted net income per share of $9.50.
- Operating cash flow was $12.0 billion.
- Capital expenditures were $4.6 billion.

EXHIBIT 27 Word Cloud Visualizing Excerpted Text in MDA Section in Form 10-Q of QXR Inc.

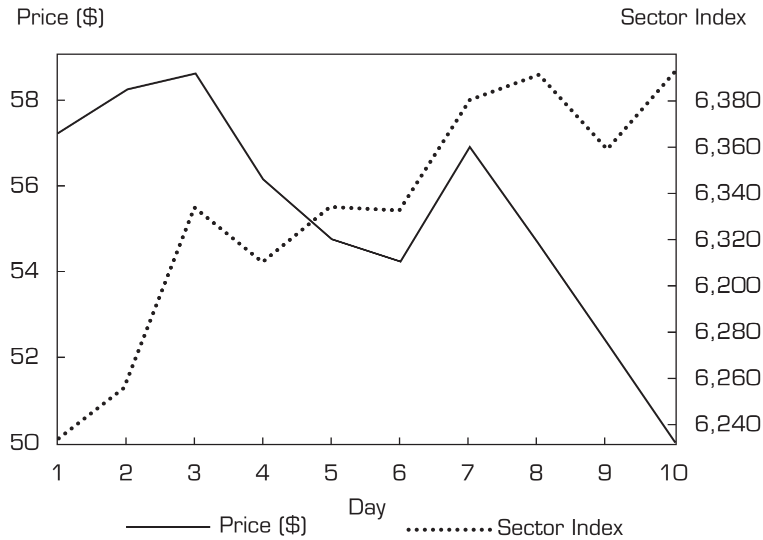
4.5. Line Chart

A **line chart** is a type of graph used to visualize ordered observations. Often a line chart is used to display the change of data series over time. Note that the frequency polygon in Exhibit 20 and the cumulative frequency distribution chart in Exhibit 21 are also line charts but used particularly in those instances for representing data frequency distributions.

Constructing a line chart is relatively straightforward: We first plot all the data points against horizontal and vertical axes and then connect the points by straight line segments. For example, to show the 10-day daily closing prices of ABC Inc. stock presented in Exhibit 5, we first construct a chart with the *x*-axis representing time (in days) and the *y*-axis representing stock price (in dollars). Next, plot each closing price as points against both axes, and then use straight line segments to join the points together, as shown in Exhibit 28.

An important benefit of a line chart is that it facilitates showing changes in the data and underlying trends in a clear and concise way. This helps to understand the current data and also helps with forecasting the data series. In Exhibit 28, for example, it is easy to spot the price changes over the first 10 trading days since ABC's initial public offering (IPO). We see that the stock price peaked on Day 3 and then traded lower. Following a partial recovery on Day 7, it declined steeply to around \$50 on Day 10. In contrast, although the

EXHIBIT 28 Daily Closing Prices of ABC Inc.'s Stock and Its Sector Index



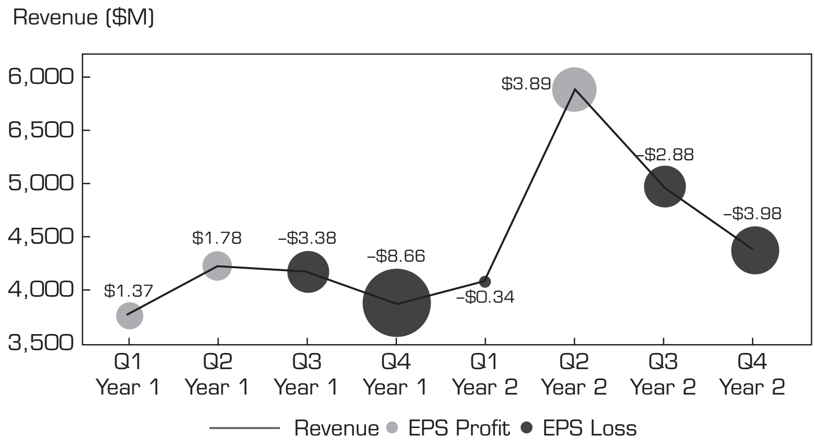
one-dimensional data array table in Exhibit 5 displays the same values as the line chart, the data table by itself does not provide a quick snapshot of changes in the data or facilitate understanding underlying trends. This is why line charts are helpful for visualization, particularly in cases of large amounts of data (i.e., hundreds, or even thousands, of data points).

A line chart is also capable of accommodating more than one set of data points, which is especially helpful for making comparisons. We can add a line to represent each group of data (e.g., a competitor's stock price or a sector index), and each line would have a distinct color or line pattern identified in a legend. For example, Exhibit 28 also includes a plot of ABC's sector index (i.e., the sector index for which ABC stock is a member, like health care or energy) over the same period. The sector index is displayed with its own distinct color or line pattern to facilitate comparison. Note also that because the sector index has a different range (approximately 6,230 to 6,390) than ABC's stock ($50 to $59 per share), we need a secondary y-axis to correctly display the sector index, which is on the right-hand side of the exhibit.

This comparison can help us understand whether ABC's stock price movement over the period is due to potential mispricing of its share issuance or instead due to industry-specific factors that also affect its competitors' stock prices. The comparison shows that over the period, the sector index moved in a nearly opposite trend versus ABC's stock price movement. This indicates that the steep decline in ABC's stock price is less likely attributable to sector-specific factors and more likely due to potential over-pricing of its IPO or to other company-specific factors.

When an observational unit (here, ABC Inc.) has more than two features (or variables) of interest, it would be useful to show the multi-dimensional data all in one chart to gain insights from a more holistic view. How can we add an additional dimension to a two-dimensional line chart? We can replace the data points with varying-sized bubbles to

EXHIBIT 29 Quarterly Revenue and EPS of ABC Incorporated



represent a third dimension of the data. Moreover, these bubbles may even be patterned, shaded or color coded to present additional information. This version of a line chart is called a **bubble line chart**.

Exhibit 7, for example, presented three types of quarterly data for ABC Inc. for use in a valuation analysis. We would like to plot two of them, revenue and earnings per share (EPS), over the two-year period. As shown in Exhibit 29, with the *x*-axis representing time (i.e., quarters) and the *y*-axis representing revenue in millions of dollars, we can plot the revenue data points against both axes to form a typical line chart. Next, each marker representing a revenue data point is replaced by a circular bubble with its size proportional to the magnitude of the EPS in the corresponding quarter. Moreover, the bubbles are shaded in a binary scheme with lightly shaded representing profits and dark shading representing losses. In this way, the bubble line chart reflects the changes for both revenue and EPS simultaneously, and it also shows whether the EPS represents a profit or a loss.

As depicted, ABC's earning were quite volatile during its initial two years as a public company. Earnings started off as a profit of $1.37/share but finished the first year with a big loss of −$8.66/share, during which time revenue experienced only small fluctuations. Furthermore, while revenues and earnings both subsequently recovered sharply—peaking in Q2 of Year 2—revenues then declined, and the company returned to significant losses (−3.98/share) by the end of Year 2.

4.6. Scatter Plot

A **scatter plot** is a type of graph for visualizing the joint variation in two numerical variables. It is a useful tool for displaying and understanding potential relationships between the variables.

A scatter plot is constructed with the *x*-axis representing one variable and the *y*-axis representing the other variable. It uses dots to indicate the values of the two variables for a particular point in time, which are plotted against the corresponding axes. Suppose an analyst is investigating potential relationships between sector index returns and returns for the broad market, such as the S&P 500 Index. Specifically, he or she is interested in the relative performance of two sectors, information technology (IT) and utilities, compared to the market index over a specific five-year period. The analyst has obtained the sector and market index returns for each month over the five years under investigation and plotted the data points in the scatter plots, shown in Exhibit 30 for IT versus the S&P 500 returns and in Exhibit 31 for utilities versus the S&P 500 returns.

Despite their relatively straightforward construction, scatter plots convey lots of valuable information. First, it is important to inspect for any potential association between the two variables. The pattern of the scatter plot may indicate no apparent relationship, a linear association, or a non-linear relationship. A scatter plot with randomly distributed data points would indicate no clear association between the two variables. However, if the data points seem to align along a straight line, then there may exist a significant relationship among the variables. A positive (negative) slope for the line of data points indicates a positive (negative) association, *meaning the variables move in the same (opposite) direction*. Furthermore, the strength of the association can be determined by how closely the data points are clustered around the line. Tight (loose) clustering signals a potentially stronger (weaker) relationship.

Examining Exhibit 30, we can see the returns of the IT sector are highly positively associated with S&P 500 Index returns because the data points are tightly clustered along a positively sloped line. Exhibit 31 tells a different story for relative performance of the utilities sector and S&P 500 index returns: The data points appear to be distributed in no discernable pattern, indicating no clear relationship among these variables. Second, observing the data points located toward the ends of each axis, which represent the maximum or minimum values, provides a quick sense of the data range. Third, assuming that a relationship among the variables is apparent, inspecting the scatter plot can help to spot extreme values (i.e., outliers). For example, an outlier data point is readily detected in Exhibit 30, as indicated by the arrow. As you will learn later in the CFA Program curriculum, finding these extreme values and handling them with appropriate measures is an important part of the financial modeling process.

EXHIBIT 30 Scatter Plot of Information Technology Sector Index Return vs. S&P 500 Index Return