CHAPTER

1

Preliminaries and Review

1.1 FINANCIAL MODELS

A model can be characterized as an artificial structure describing the relationships among variables or factors. Practically all of the methodology in this book is geared toward the development and implementation of financial models to solve financial problems. For example, valuation models provide a foundation for investment decision-making and models describing stochastic processes provide an important tool to account for risk in decision-making.

The use of models is important in finance because "real world" conditions that underlie financial decisions are frequently extraordinarily complicated. Financial decision-makers frequently use existing models or construct new ones that relate to the types of decisions they wish to make. Models proposing decisions that ought to be made are called normative models.[1]

The purpose of models is to simulate or behave like real financial situations. When constructing financial models, analysts exclude the "real world" conditions that seem to have little or no effect on the outcomes of their decisions, concentrating on those factors that are most relevant to their situations. In some instances, analysts may have to make unrealistic assumptions in order to simplify their models and make them easier to analyze. After simple models have been constructed with what may be unrealistic assumptions, they can be modified to match more closely "real world" situations. A good financial model is one that accounts for the major factors that will affect the financial decision (a good model is complete and accurate), is simple enough for its use to be practical (inexpensive to construct and easy to understand), and can be used to predict actual outcomes. A model is not likely to be useful if it is not able to project an outcome with an acceptable degree of accuracy. Completeness and simplicity may directly conflict with one another. The financial analyst must determine the appropriate trade-off between completeness and simplicity in the model he wishes to construct.

In finance, mathematical models are usually the easiest to develop, manipulate, and modify. These models are usually adaptable to computers and electronic spreadsheets. Mathematical models are obviously most useful for those comfortable with math; the primary purpose of this book is to provide a foundation for improving the quantitative preparation of the less mathematically oriented analyst. Other models used in finance include those based on graphs and those involving simulations. However, these models are often based on or closely related to mathematical models.

P.M. Knopf & J.L. Teall: Risk Neutral Pricing and Financial Mathematics: A Primer.
DOI: http://dx.doi.org/10.1016/B978-0-12-801534-6.00001-1

The concepts of *market efficiency* and *arbitrage* are essential to the development of many financial models. Market efficiency is the condition in which security prices fully reflect all available information. Such efficiency is more likely to exist when wealth-maximizing market participants can instantaneously and costlessly execute transactions as information is revealed. Transactions costs, irrationality, and poor execution systems reduce efficiency. Arbitrage, in its simplest scenario, is the simultaneous purchase and sale of the same asset, or more generally, the nearly simultaneous purchase and sale of assets generating nearly identical cash flow structures. In either case, the arbitrageur seeks to produce a profit by purchasing at a price that is less than the selling price. Proceeds of the sales are used to finance purchases such that the portfolio of transactions is self-financing, and that over time, no additional capital is devoted to or lost from the portfolio. Thus, the portfolio is assured a non-negative profit at each time period. The arbitrage process is riskless if purchase and sale prices are known at the times they are initiated. Arbitrageurs frequently seek to profit from market inefficiencies. The existence of arbitrage profits is inconsistent with market efficiency.

1.2 FINANCIAL SECURITIES AND INSTRUMENTS

A security is a tradable claim on assets. Real assets contribute to the productive capacity of the economy; securities are financial assets that represent claims on real assets or other securities. Most securities are marketable to the general public, meaning that they can be sold or assigned to other investors in the open marketplace. Some of the more common types of securities and tradable instruments are briefly introduced in the following:

1. *Debt securities*: Denote creditorship of an individual, firm or other institution. They typically involve payments of a fixed series of interest (often known as *coupon payments*) or amounts towards principal along with principal repayment (often known as *face value*). Examples include:
   - *Bonds*: Long-term debt securities issued by corporations, governments, or other institutions. Bonds are normally of the coupon variety (they make periodic interest payments on the principal) or *pure discount* (they are *zero coupon instruments* that are sold at a discount from *face value*, the bond's final maturity value).
   - *Treasury securities:* Debt securities issued by the Treasury of the United States federal government. They are often considered to be practically free of default risk.
2. *Equity securities (stock)*: Denote ownership in a business or corporation. They typically permit for dividend payments if the firm's debt obligations have been satisfied.
3. *Derivative securities*: Have payoff functions derived from the values of other securities, rates, or indices. Some of the more common derivative securities are:
   - *Options*: Securities that grant their owners rights to buy (call) or sell (put) an underlying asset or security at a specific price (exercise price) on or before its expiration date.
   - *Forward* and *futures contracts*: Instruments that oblige their participants to either purchase or sell a given asset or security at a specified price (settlement price) on the future settlement date of that contract. A *long* position obligates the investor to purchase the given asset on the settlement date of the contract and a *short* position obligates the investor to sell the given asset on the settlement date of the contract.
   - *Swaps*: Provide for the exchange of cash flows associated with one asset, rate, or index for the cash flows associated with another asset, rate, or index.

4. *Commodities*: Contracts, including futures and options on physical commodities such as oil, metals, corn, etc. Commodities are traded in *spot markets*, where the exchange of assets and money occurs at the time of the transaction or in forward and futures markets.
5. *Currencies*: Exchange rates denote the number of units of one currency that must be given up for one unit of a second currency. Exchange transactions can occur in either spot or forward markets. As with commodities, in the spot market, the exchange of one currency for another occurs when the agreement is made. In a forward market transaction, the actual exchange of one currency for another actually occurs at a date later than that of the agreement. Spot and forward contract participants take one position in each of two currencies:
   - *Long*: An investor has a "long" position in that currency that he will accept at the later date.
   - *Short*: An investor has a "short" position in that currency that he must deliver in the transaction.
6. *Indices*: Contracts pegged to measures of market performance such as the Dow Jones Industrials Average or the S&P 500 Index. These are frequently futures contracts on portfolios structured to perform exactly as the indices for which they are named. Index traders also trade options on these futures contracts.

This list of security types is far from comprehensive; it only reflects some of those instruments that will be emphasized in this book. In addition, most of the instrument types will have many different variations.

1.3 REVIEW OF MATRICES AND MATRIX ARITHMETIC

A *matrix* is simply an ordered rectangular array of numbers. A matrix is an entity that enables one to represent a series of numbers as a single object, thereby providing for convenient systematic methods for completing large numbers of repetitive computations. Such objects are essential for the management of large data structures. Rules of matrix arithmetic and other matrix operations are often similar to rules of ordinary arithmetic and other operations, but they are not always identical. In this text, matrices will usually be denoted with bold uppercase letters. When the matrix has only one row or one column, bold lowercase letters will be used for identification. The following are examples of matrices:

$$\mathbf{A} = \begin{bmatrix} 4 & 2 & 6 \\ 3 & 7 & 4 \\ 8 & -5 & 9 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 2 & -3 \\ 3/4 & -1/2 \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} 1 \\ 5 \\ 7 \end{bmatrix} \quad \mathbf{d} = [4]$$

The dimensions of a matrix are given by the ordered pair $m \times n$, where m is the number of rows and n is the number of columns in the matrix. The matrix is said to be of *order* $m \times n$ where, by convention, the number of rows is listed first. Thus, $\mathbf{A}$ is 3×3, $\mathbf{B}$ is 2×2, $\mathbf{c}$ is 3×1, and $\mathbf{d}$ is 1×1. Each number in a matrix is referred to as an element. The symbol $a_{i,j}$ denotes the element in Row i and Column j of Matrix $\mathbf{A}$, $b_{i,j}$ denotes the element in Row i and Column j of Matrix $\mathbf{B}$, and so on. Thus, $a_{3,2}$ is -5 and $c_{2,1} = 5$.

There are specific terms denoting various types of matrices. Each of these particular types of matrices has useful applications and unique properties for working with. For example, a *vector* is a matrix with either only one row or one column. Thus, the dimensions of a vector are $1 \times n$ or $m \times 1$. Matrix **c** above is a column vector; it is of order 3×1. A $1 \times n$ matrix is a row vector with n elements. The column vector has one column and the row vector has one row. A *scalar* is a 1×1 matrix with exactly one entry, which means that a scalar is simply a number. Matrix **d** is a scalar, which we normally write as simply the number 4. A *square matrix* has the same number of rows and columns ($m = n$). Matrix **A** is square and of order 3. The set of elements extending from the upper leftmost corner to the lower rightmost corner in a square matrix is said to be on the *principal diagonal*. For a square matrix **A**, each of these elements are those of the form $a_{i,j}$, $i = j$. Principal diagonal elements of Square Matrix **A** in the example above are $a_{1,1} = 4$, $a_{2,2} = 7$, and $a_{3,3} = 9$. Matrices **B** and **d** are also square matrices.

A *symmetric matrix* is a square matrix where $c_{i,j}$ equals $c_{j,i}$ for all i and j. This is equivalent to the condition kth row equals the kth column for every k. Scalar **d** and matrices **H**, **I**, and **J** below are all symmetric matrices. A *diagonal matrix* is a symmetric matrix whose elements off the principal diagonal are zero, where the *principal diagonal* contains the series of elements where $i = j$. Scalar **d** and Matrices **H** and **I** below are all diagonal matrices. An *identity* or *unit* matrix is a diagonal matrix consisting of ones along the principal diagonal. Matrix **I** below is the 3×3 identity matrix:

$$\mathbf{H} = \begin{bmatrix} 13 & 0 & 0 \\ 0 & 11 & 0 \\ 0 & 0 & 10 \end{bmatrix} \quad \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{J} = \begin{bmatrix} 1 & 7 & 2 \\ 7 & 5 & 0 \\ 2 & 0 & 4 \end{bmatrix}$$

1.3.1 Matrix Arithmetic

Matrix arithmetic provides for standard rules of operation just as conventional arithmetic. Matrices can be added or subtracted if their dimensions are identical. Matrices **A** and **B** add to **C** if $a_{i,j} + b_{i,j} = c_{i,j}$ for all i and j:

$$\underset{\mathbf{A}}{\begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix}} \underset{+}{+} \underset{\mathbf{B}}{\begin{bmatrix} b_{1,1} & b_{1,2} & \dots & b_{1,n} \\ b_{2,1} & b_{2,2} & \dots & b_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{m,1} & b_{m,2} & \dots & b_{m,n} \end{bmatrix}} \underset{=}{=} \underset{\mathbf{C}}{\begin{bmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,n} \\ c_{2,1} & c_{2,2} & \dots & c_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ c_{m,1} & c_{m,2} & \dots & c_{m,n} \end{bmatrix}}$$

Note that each of the three matrices is of dimension 3×3 and that each of the elements in Matrix **C** is the sum of corresponding elements in Matrices **A** and **B**. The process of subtracting matrices is similar, where $d_{i,j} - e_{i,j} = f_{i,j}$ for $\mathbf{D} - \mathbf{E} = \mathbf{F}$:

$$\underset{\mathbf{D}}{\begin{bmatrix} d_{1,1} & d_{1,2} & \dots & d_{1,n} \\ d_{2,1} & d_{2,2} & \dots & d_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ d_{m,1} & d_{m,2} & \dots & d_{m,n} \end{bmatrix}} \underset{-}{-} \underset{\mathbf{E}}{\begin{bmatrix} e_{1,1} & e_{1,2} & \dots & e_{1,n} \\ e_{2,1} & e_{2,2} & \dots & e_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ e_{m,1} & e_{m,2} & \dots & e_{m,n} \end{bmatrix}} \underset{=}{=} \underset{\mathbf{F}}{\begin{bmatrix} f_{1,1} & f_{1,2} & \dots & f_{1,n} \\ f_{2,1} & f_{2,2} & \dots & f_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ f_{m,1} & f_{m,2} & \dots & f_{m,n} \end{bmatrix}}$$

Now consider a third matrix operation. The *transpose* $\mathbf{A}^{\mathrm{T}}$ of Matrix **A** is obtained by interchanging the rows and columns of Matrix **A**. Each $a_{i,j}$ becomes $a_{j,i}$. The following represent Matrix **A** and its transpose $\mathbf{A}^{\mathrm{T}}$:

$$\underset{\mathbf{A}}{\begin{bmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n} \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \ldots & a_{m,n} \end{bmatrix}}, \quad \underset{\mathbf{A}^{\mathrm{T}}}{\begin{bmatrix} a_{1,1} & a_{2,1} & \ldots & a_{m,1} \\ a_{1,2} & a_{2,2} & \ldots & a_{m,2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1,n} & a_{2,n} & \ldots & a_{m,n} \end{bmatrix}}$$

Two matrices **A** and **B** can be multiplied to obtain the product $\mathbf{AB} = \mathbf{C}$ if the number of columns in the first Matrix **A** equals the number of rows **B** in the second.[2] If Matrix **A** is of dimension $m \times n$ and Matrix **B** is of dimension $n \times q$, the dimensions of the product Matrix **C** will be $m \times q$. Each element $c_{i,k}$ of Matrix **C** is determined by the following sum:

$$c_{i,k} = \sum_{j=1}^{n} a_{i,j} b_{j,k}$$

$$\underset{\mathbf{A}}{\begin{bmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n} \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \ldots & a_{m,n} \end{bmatrix}} \underset{\times}{\times} \underset{\mathbf{B}}{\begin{bmatrix} b_{1,1} & b_{1,2} & \ldots & b_{1,q} \\ b_{2,1} & b_{2,2} & \ldots & b_{2,q} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n,1} & b_{n,2} & \ldots & b_{n,q} \end{bmatrix}}$$

$$= \underset{\mathbf{A} \times \mathbf{B}}{\begin{bmatrix} \sum_{j=1}^{n} a_{1,j} b_{j,1} & \sum_{j=1}^{n} a_{1,j} b_{j,2} & \ldots & \sum_{j=1}^{n} a_{1,j} b_{j,q} \\ \sum_{j=1}^{n} a_{2,j} b_{j,1} & \sum_{j=1}^{n} a_{2,j} b_{j,2} & \ldots & \sum_{j=1}^{n} a_{2,j} b_{j,q} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{j=1}^{n} a_{m,j} b_{j,1} & \sum_{j=1}^{n} a_{m,j} b_{j,m} & \ldots & \sum_{j=1}^{n} a_{m,j} b_{j,q} \end{bmatrix}} \underset{=}{=} \underset{\mathbf{C}}{\begin{bmatrix} c_{1,1} & c_{1,2} & \ldots & c_{1,q} \\ c_{2,1} & c_{2,2} & \ldots & c_{2,q} \\ \vdots & \vdots & \vdots & \vdots \\ c_{m,1} & c_{m,2} & \ldots & c_{m,q} \end{bmatrix}}$$

Notice that the number of columns (n) in Matrix **A** equals the number of rows in Matrix **B**. Also note that the number of rows in Matrix **C** equals the number of rows in Matrix **A**; the number of columns in **C** equals the number of columns in Matrix **B**. One additional detail on matrix multiplication is that scalar multiplication is the product of a real number c with a matrix **A**:

$$c\mathbf{A} = \begin{bmatrix} ca_{1,1} & ca_{1,2} & \ldots & ca_{1,n} \\ ca_{2,1} & ca_{2,2} & \ldots & ca_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ ca_{m,1} & ca_{m,2} & \ldots & ca_{m,n} \end{bmatrix}$$

Matrix Arithmetic Illustration:

Consider the following matrices **A** and **B** below:

$$\mathbf{A} = \begin{bmatrix} 3 & 0 \\ -2 & -1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 5 & 2 \\ -6 & 4 \end{bmatrix}$$

We find $\mathbf{A}^T$, 4**A**, **A** + **B**, **AB**, and **BA** as follows:

$$\mathbf{A}^T = \begin{bmatrix} 3 & -2 \\ 0 & -1 \end{bmatrix}, \quad 4\mathbf{A} = \begin{bmatrix} 4(3) & 4(0) \\ 4(-2) & 4(-1) \end{bmatrix} = \begin{bmatrix} 12 & 0 \\ -8 & -4 \end{bmatrix}$$

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 3+5 & 0+2 \\ -2-6 & -1+4 \end{bmatrix} = \begin{bmatrix} 8 & 2 \\ -8 & 3 \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} 3(5)+0(-6) & 3(2)+0(4) \\ (-2)(5)+(-1)(-6) & -2(2)+(-1)(4) \end{bmatrix} = \begin{bmatrix} 15 & 6 \\ -4 & -8 \end{bmatrix}$$

$$\mathbf{BA} = \begin{bmatrix} 5(3)+2(-2) & 5(0)+2(-1) \\ (-6)(3)+4(-2) & -6(0)+4(-1) \end{bmatrix} = \begin{bmatrix} 11 & -2 \\ -26 & -4 \end{bmatrix}$$

1.3.1.1 Matrix Arithmetic Properties

It is useful to note that matrices have certain algebraic properties that are similar to the algebraic properties of real numbers. Here are a few of their properties:

1. **A** + **B** = **B** + **A** (commutative property of addition)
2. **A**(**B** + **C**) = **AB** + **AC** (distributive property)
3. **AI** = **IA** = **A** where **I** is the identity matrix
4. $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$

However, it is important to observe that, unlike real numbers, the commutative property of multiplication does not hold for matrices; that is, in general, $\mathbf{AB} \neq \mathbf{BA}$.

1.3.1.2 The Inverse Matrix

An *inverse* Matrix $\mathbf{A}^{-1}$ exists for the square Matrix **A** if the products $\mathbf{AA}^{-1}$ or $\mathbf{A}^{-1}\mathbf{A}$ equal the identity Matrix **I**:

$$\mathbf{AA}^{-1} = \mathbf{I}$$
$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

One means for finding the inverse Matrix $\mathbf{A}^{-1}$ for Matrix **A** is through the use of a process called the *Gauss–Jordan method*.

ILLUSTRATION: THE GAUSS–JORDAN METHOD

An *inverse* Matrix $\mathbf{A}^{-1}$ exists for the square Matrix $\mathbf{A}$ if the product $\mathbf{A}^{-1}\mathbf{A}$ or $\mathbf{A}\mathbf{A}^{-1}$ equals the identity Matrix $\mathbf{I}$. Consider the following product:

$$\textbf{A.}\quad \begin{bmatrix} 2 & 4 \\ 8 & 1 \end{bmatrix} \begin{bmatrix} \frac{-1}{30} & \frac{2}{15} \\ \frac{4}{15} & \frac{-1}{15} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{A} \qquad \mathbf{A}^{-1} \qquad = \qquad \mathbf{I}$$

To construct $\mathbf{A}^{-1}$ given a square matrix $\mathbf{A}$, we will use the Gauss–Jordan method. We illustrate the method with the example above. First, augment $\mathbf{A}$ with the 2×2 identity matrix as follows:

$$\textbf{B.}\quad \left[\begin{array}{cc:cc} 2 & 4 & 1 & 0 \\ 8 & 1 & 0 & 1 \end{array}\right]$$

For the sake of convenience, call the above augmented Matrix $\mathbf{B}$. Now, a series of *elementary row operations* (involves addition, subtraction, and multiplication of rows, as described below) will be performed such that the identity matrix replaces the original Matrix $\mathbf{A}$ (on the left side). The right-side elements will comprise the inverse Matrix $\mathbf{A}^{-1}$. Thus, in our final augmented matrix, we will have ones along the principal diagonal on the left side and zeros elsewhere; the right side of the matrix will comprise the inverse of $\mathbf{A}$. Allowable elementary row operations include the following:

1. Multiply a given row by any constant. Each element in the row must be multiplied by the same constant.
2. Add a given row to any other row in the matrix. Each element in a row is added to the corresponding element in the same column of another row.
3. Subtract a given row from any other row in the matrix. Each element in a row is subtracted from the corresponding element in the same column of another row.
4. Any combination of the above. For example, a row may be multiplied by a constant before it is subtracted from another row.

Our first row operation will serve to replace the upper left corner value with a one. We multiply Row 1 in $\mathbf{B}$ by .5:

$$\mathbf{B} = \left[\begin{array}{cc:cc} \mathbf{2} & 4 & 1 & 0 \\ 8 & 1 & 0 & 1 \end{array}\right] \xrightarrow{(\text{row1}) \times .5} \left[\begin{array}{cc:cc} 1 & 2 & .5 & 0 \\ 8 & 1 & 0 & 1 \end{array}\right] = \mathbf{C}$$

Now we obtain a zero in the lower left corner by multiplying Row 2 in **C** by 1/8 and subtracting the result from Row 1 of **C** as follows:

$$\mathbf{C} = \begin{bmatrix} 1 & 2 & \vdots & .5 & 0 \\ \mathbf{8} & 1 & \vdots & 0 & 1 \end{bmatrix} \xrightarrow{\text{row1} - 1/8(\text{row2})} \begin{bmatrix} 1 & 2 & \vdots & .5 & 0 \\ 0 & \frac{15}{8} & \vdots & .5 & \frac{-1}{8} \end{bmatrix} = \mathbf{D}$$

Next, we obtain a 1 in the lower right corner of the left side of the matrix by multiplying Row 2 of matrix **D** by 8/15:

$$\mathbf{D} = \begin{bmatrix} 1 & 2 & \vdots & .5 & 0 \\ 0 & \mathbf{\frac{15}{8}} & \vdots & .5 & \frac{-1}{8} \end{bmatrix} \xrightarrow{(\text{row2}) \times \frac{8}{15}} \begin{bmatrix} 1 & 2 & \vdots & .5 & 0 \\ 0 & 1 & \vdots & \frac{4}{15} & \frac{-1}{15} \end{bmatrix} = \mathbf{E}$$

We obtain a zero in the upper right corner of the left-side matrix by multiplying Row 2 of matrix **E** above by 2 and subtracting from Row 1 in **E**:

$$\mathbf{E} = \begin{bmatrix} 1 & \mathbf{2} & \vdots & .5 & 0 \\ 0 & 1 & \vdots & \frac{4}{15} & \frac{-1}{15} \end{bmatrix} \xrightarrow{\text{row1} - (\text{row2}) \times 2} \begin{bmatrix} 1 & 0 & \vdots & \frac{-1}{30} & \frac{2}{15} \\ 0 & 1 & \vdots & \frac{4}{15} & \frac{-1}{15} \end{bmatrix} = \mathbf{F}$$

The left side of augmented Matrix **F** is the identity matrix; the right side of **F** is $\mathbf{A}^{-1}$.

ILLUSTRATION: SOLVING SYSTEMS OF EQUATIONS

Matrices can be very useful in arranging systems of equations. Consider, for example, the following system of equations:

$$.05x_1 + .12x_2 = .05$$
$$.10x_1 + .30x_2 = .08$$

This system of equations can be represented as follows:

$$\begin{array}{ccccc} \begin{bmatrix} .05 & .12 \\ .10 & .30 \end{bmatrix} & & \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} & = & \begin{bmatrix} .05 \\ .08 \end{bmatrix} \\ \mathbf{C} & \times & \mathbf{x} & = & \mathbf{s} \end{array}$$

Thus, we can express this system of equations as the matrix equation $\mathbf{Cx} = \mathbf{s}$, where in general **C** is a given $n \times n$ matrix, **s** is a given $n \times 1$ column vector, and **x** is the unknown $n \times 1$ column vector for which we wish to solve. In ordinary algebra, if we had the real-valued equation $Cx = s$, we would solve for s by dividing both sides of the equation by C, which is equivalent to multiplying both sides of the equation by the inverse of C. Here we show the algebra, so that we see that this process with real numbers is essentially equivalent for the process with matrices:

$$Cx = s,\ C^{-1}Cx = C^{-1}s,\ 1(x) = C^{-1}s,\ x = C^{-1}s$$

With matrices, the process is:

$$\mathbf{Cx} = \mathbf{s},\ \mathbf{C}^{-1}\mathbf{Cx} = \mathbf{C}^{-1}\mathbf{s},\ \mathbf{Ix} = \mathbf{C}^{-1}\mathbf{s},\ \mathbf{x} = \mathbf{C}^{-1}\mathbf{s}$$

Of course, in ordinary algebra, it is trivial to find the inverse of a number C, which is simply its reciprocal $1/C$. To find the inverse of a matrix $\mathbf{C}$, we use the Gauss–Jordan method described above. We begin by augmenting the matrix $\mathbf{C}$ by placing its corresponding identity matrix $\mathbf{I}$ immediately to its right:

A. $$\begin{bmatrix} .05 & .12 & \vdots & 1 & 0 \\ .10 & .30 & \vdots & 0 & 1 \end{bmatrix}$$

We will reduce this matrix using the allowable elementary row operations described earlier to the form with the identity matrix $\mathbf{I}$ on the left replacing $\mathbf{C}$, and to the right will be the inverse of $\mathbf{C}$:

B. $$\begin{bmatrix} 1 & 2.4 & \vdots & 20 & 0 \\ 0 & .6 & \vdots & -20 & 10 \end{bmatrix} \quad \begin{array}{l} \text{Row } B1 = A1 \cdot 20 \\ \text{Row } B2 = (10 \cdot A2) - B1 \end{array}$$

C. $$\begin{bmatrix} 1 & 0 & \vdots & 100 & -40 \\ 0 & 1 & \vdots & \dfrac{-100}{3} & \dfrac{50}{3} \end{bmatrix} \quad \begin{array}{l} \text{Row } C1 = B1 - (2.4 \cdot C2) \\ \text{Row } C2 = B2 \cdot 5/3 \end{array}$$

$$\mathbf{I} \qquad \mathbf{C}^{-1}$$

Thus, we obtain Vector $\mathbf{x}$ with the following product:

D. $$\underset{\mathbf{x}}{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}} = \underset{\mathbf{C}^{-1}}{\begin{bmatrix} 100 & -40 \\ \dfrac{-100}{3} & \dfrac{50}{3} \end{bmatrix}} \times \underset{\mathbf{s}}{\begin{bmatrix} .05 \\ .08 \end{bmatrix}} = \begin{bmatrix} 1.8 \\ \dfrac{-1}{3} \end{bmatrix}$$

Thus, we find that $x_1 = 1.8$ and $x_2 = -1/3$.

1.3.2 Vector Spaces, Spanning, and Linear Dependence

$\mathbb{R}^n$ is defined as the set of all vectors (may be represented as a column or row vectors) with n real-valued entries or coordinates. The row vector $\mathbf{x}^{\mathrm{T}} = (x_1, x_2, \ldots, x_n)$ or column vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ can be regarded as a point in the n-dimensional space $\mathbb{R}^n$ and x_i is the ith coordinate of the point (vector) $\mathbf{x}$.

The set $\mathbb{R}^n$ with the operations of vector addition and scalar multiplication (discussed earlier) makes $\mathbb{R}^n$ an n-dimensional vector space. A *linear combination* of vectors is accomplished through either or both of the following:

- Multiplication of any vector by a scalar (real number)
- Addition of any combination of vectors either before or after multiplication by scalars

1.3.2.1 *Linear Dependence and Linear Independence*

If a vector in $\mathbb{R}^n$ can be expressed as a linear combination of a set of other vectors in $\mathbb{R}^n$, then that set of vectors including the first is said to be *linearly dependent*. Suppose we are given a set of m vectors: $\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m\}$ with each vector $\mathbf{x}_i$ in $\mathbb{R}^n$. An equivalent definition of linear dependence of the set of vectors $\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m\}$ is that there exists m scalars: $\alpha_1,\alpha_2,\ldots,\alpha_m$ so that:

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \alpha_3\mathbf{x}_3 + \cdots + \alpha_m\mathbf{x}_m = \mathbf{0}$$

where at least one of the scalars α_i is non-zero and $\mathbf{0} = (0,0,\ldots,0)$ or $\mathbf{0} = (0,0,\ldots,0)^T$ depending upon whether the vectors $\mathbf{x}_i$ are expressed as row or column vectors. We note that $\mathbf{0}$ is called the zero vector. The set of vectors $\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m\}$ is said to be linearly independent when the only set of scalars $\{\alpha_1,\alpha_2,\ldots,\alpha_m\}$ that satisfy the equation above is when $\alpha_i = 0$ for all $i = 1,2,\ldots,m$. When the set of vectors $\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m\}$ is linearly independent, then no vector in this set can be expressed as a linear combination of the other vectors in the set. If we denote the $n \times m$ matrix $\mathbf{X} = [\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m]$ and the $m \times 1$ column vector of scalars $\boldsymbol{\alpha} = (\alpha_1,\alpha_2,\ldots,\alpha_m)^T$, then we can express the above equation as the matrix equation $\mathbf{X}\boldsymbol{\alpha} = \mathbf{0}$, where $\mathbf{0} = (0,0,\ldots,0)^T$ is the $n \times 1$ column zero vector.

ILLUSTRATIONS: LINEAR DEPENDENCE AND INDEPENDENCE

Consider the following set $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ of three vectors in $\mathbb{R}^3$:

$$\underset{\mathbf{x}_1}{\begin{bmatrix}3\\1\\9\end{bmatrix}} \quad \underset{\mathbf{x}_2}{\begin{bmatrix}5\\5\\15\end{bmatrix}} \quad \underset{\mathbf{x}_3}{\begin{bmatrix}1\\2\\3\end{bmatrix}} \qquad \alpha_1\mathbf{x_1} + \alpha_2\mathbf{x_2} + \alpha_3\mathbf{x_3} = [\mathbf{0}]$$

We will determine whether this set is linearly independent. Let vector $\boldsymbol{\alpha}$ be $[\alpha_1,\ \alpha_2,\ \alpha_3]^T$ and Matrix $\mathbf{X}$ be $[\mathbf{x}_1,\ \mathbf{x}_2,\ \mathbf{x}_3]$. We determine that vector set $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \mathbf{x}_3\}$ is linearly dependent by demonstrating that there exists a vector $\boldsymbol{\alpha}$ that produces $\mathbf{X}\boldsymbol{\alpha} = [\mathbf{0}]$. By inspection, we find that $\boldsymbol{\alpha} = [1,\ -1,\ 2]^T$ is one such vector. Thus, the set $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \mathbf{x}_3\}$ is linearly dependent. Also note that any one of these three vectors is a linear combination of the other two.

Vector set $\{\mathbf{y}_1,\ \mathbf{y}_2,\ \mathbf{y}_3\}$ below is linearly independent because the only vector satisfying $\boldsymbol{\alpha}^T\mathbf{Y} = [\mathbf{0}]$ is $\boldsymbol{\alpha} = [0,\ 0,\ 0]^T$:[3]

$$\underset{\mathbf{y}_1}{\begin{bmatrix}3\\1\\9\end{bmatrix}} \quad \underset{\mathbf{y}_2}{\begin{bmatrix}5\\5\\15\end{bmatrix}} \quad \underset{\mathbf{y}_3}{\begin{bmatrix}1\\2\\4\end{bmatrix}} \qquad \alpha_1\mathbf{y_1} + \alpha_2\mathbf{y_2} + \alpha_3\mathbf{y_3} = [\mathbf{0}]$$

Furthermore, no vector in set $\{\mathbf{y}_1,\ \mathbf{y}_2,\ \mathbf{y}_3\}$ can be defined as a linear combination of the other vectors in set $\{\mathbf{Y}\}$. Thus, this set is linearly independent. This means that it is impossible to express any one of the vectors as a linear combination of the other two vectors.

1.3.2.2 *Spanning the Vector Space and the Basis*

A set of m vectors $\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m\}$, where each vector $\mathbf{x}_i$ is an n-dimensional vector in $\mathbb{R}^n$, is said to *span* the n-dimensional vector space $\mathbb{R}^n$ if any vector in $\mathbb{R}^n$ can be expressed as a linear combination of the vectors $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m$. In other words, for every vector $\mathbf{v}$ in $\mathbb{R}^n$, there exist scalars $\alpha_1,\alpha_2,\ldots,\alpha_m$ such that $\mathbf{v} = \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \ldots + \alpha_m\mathbf{x}_m$.

If a set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ is both linearly independent and spans the n-dimensional space $\mathbb{R}^n$, then that set of vectors is called a *basis* for the vector space $\mathbb{R}^n$. However, any basis for $\mathbb{R}^n$ must consist of exactly n vectors. This is because for a set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ in $\mathbb{R}^n$, if $m < n$, then there are not enough vectors to span $\mathbb{R}^n$. On the other hand, if $m > n$, then it is possible for the set of vectors to span $\mathbb{R}^n$, but there will be too many such vectors for the set to be linearly independent. Thus, any set of $m = n$ linearly independent vectors in $\mathbb{R}^n$ will form a basis for $\mathbb{R}^n$ since any such set will also always span $\mathbb{R}^n$.

ILLUSTRATION: SPANNING THE VECTOR SPACE AND THE BASIS

We return to our illustration above with our linearly independent vector set $\{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3\}$:

$$\underset{\mathbf{y}_1}{\begin{bmatrix} 3 \\ 1 \\ 9 \end{bmatrix}} \quad \underset{\mathbf{y}_2}{\begin{bmatrix} 5 \\ 5 \\ 15 \end{bmatrix}} \quad \underset{\mathbf{y}_3}{\begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}}$$

Since this set is linearly independent, it will form a basis for $\mathbb{R}^3$ if it also spans the three-dimensional space. We will demonstrate that any vector $\mathbf{v}$ in $\mathbb{R}^3$ is a linear combination of $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$, thereby demonstrating that vectors $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$ span $\mathbb{R}^3$:

$$\mathbf{v} = \alpha_1 \begin{bmatrix} 3 \\ 1 \\ 9 \end{bmatrix} + \alpha_2 \begin{bmatrix} 5 \\ 5 \\ 15 \end{bmatrix} + \alpha_3 \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

To obtain numerical values for α_1, α_2, and α_3, we combine vectors $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$ into a 3×3 matrix, then invert and multiply by $\mathbf{v}$ as follows:[4]

$$\underset{\mathbf{Y}}{\begin{bmatrix} 3 & 5 & 1 \\ 1 & 5 & 2 \\ 9 & 15 & 4 \end{bmatrix}} \underset{\boldsymbol{\alpha}}{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}} \underset{=}{=} \underset{\mathbf{v}}{\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}}$$

$$\underset{\boldsymbol{\alpha}}{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}} \underset{=}{=} \underset{\mathbf{Y}^{-1}}{\begin{bmatrix} -1 & -.5 & .5 \\ 1.4 & .3 & -.5 \\ -3 & 0 & 1 \end{bmatrix}} \underset{\times}{} \underset{\mathbf{v}}{\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}}$$

Thus, we can replicate any vector $\mathbf{v}$ with a linear combination of vectors $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$ and some vector $\boldsymbol{\alpha}$. For example, if $\mathbf{v} = [6\ 3\ 1]^T$, then we obtain $\boldsymbol{\alpha}$ as follows:

$$\underset{\boldsymbol{\alpha}}{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}} \underset{=}{=} \underset{\mathbf{Y}^{-1}}{\begin{bmatrix} -1 & -.5 & .5 \\ 1.4 & .3 & -.5 \\ -3 & 0 & 1 \end{bmatrix}} \underset{\times}{} \underset{\mathbf{v}}{\begin{bmatrix} 6 \\ 3 \\ 1 \end{bmatrix}} = \begin{bmatrix} -7 \\ 8.8 \\ -17 \end{bmatrix}$$

$$\mathbf{v} = -7 \begin{bmatrix} 3 \\ 1 \\ 9 \end{bmatrix} + 8.8 \begin{bmatrix} 5 \\ 5 \\ 15 \end{bmatrix} - 17 \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

As long as we can invert 3×3 matrix $\mathbf{Y}$, we can replicate any vector in $\mathbb{R}^3$ with some linear combination of vectors $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$ from which coefficients are obtained from vector $\boldsymbol{\alpha}$.

In a sense, when an $n+1^{\text{st}}$ vector is linearly dependent on a set of n other $n\times1$ vectors, the characteristics or information in the n other $n\times1$ vectors can be used to replicate the information in the $n+1^{\text{st}}$ vector. In a financial sense where elements in a vector represent security payoffs over time or across potential outcomes, the payoff structure of the $n+1^{\text{st}}$ security can be replicated with a portfolio comprising the n other $n\times1$ security vectors. When a set of n payoff vectors spans the n-dimensional outcome or time space, the payoff structure for any other security or portfolio in the same outcome or time space can be replicated with the payoff vectors of the n-security basis. Securities or portfolios whose payoff vectors can be replicated by portfolios of other securities must sell for the same price as those portfolios; otherwise, the *law of one price* is violated.[5]

1.4 REVIEW OF DIFFERENTIAL CALCULUS

The derivative and the integral are the two most essential concepts from calculus. The derivative from calculus can be used to determine rates of change or slopes. They are also useful for finding function maxima and minima. For those functions whose slopes are changing, the derivative is equal to the instantaneous rate of change; that is, the change in y induced by the "tiniest" change in x. Assume that y is given as a function of variable x. If x were to increase by a small (infinitesimal—that is, approaching, though not quite equal to zero) amount Δx, by how much would y change? This rate of change is given by the derivative of y with respect to x, which is defined as follows:

$$\frac{\mathrm{d}y}{\mathrm{d}x}=f'(x)=\lim_{\Delta x\to0}\frac{f(x+\Delta x)-f(x)}{\Delta x} \tag{1.1}$$

Consider Figure 1.1, which plots the function $y=2x-x^2$. Using Eq. (1.1), we will find that $\mathrm{d}y/\mathrm{d}x$, the slope of our function is calculated by:

$$\frac{\mathrm{d}y}{\mathrm{d}x}=f'(x)=\lim_{\Delta x\to0}\frac{f(x+\Delta x)-f(x)}{\Delta x}=\lim_{\Delta x\to0}\frac{2(x+\Delta x)-(x+\Delta x)^2-2x+x^2}{\Delta x}$$

$$=\lim_{\Delta x\to0}\frac{2x+2\Delta x-x^2-(\Delta x)^2-2x\Delta x-2x+x^2}{\Delta x}=\lim_{\Delta x\to0}\frac{2\Delta x-(\Delta x)^2-2x\Delta x}{\Delta x}=\lim_{\Delta x\to0}(2-\Delta x-2x)$$

$$=2-2x$$

On Figure 1.1, suppose that we start from point $(x_0,\ y_0)=(0.2,\ 0.36)$. If the change in x were $\Delta x=.8$, the change in y would be $\Delta y=(1-.36)=.64$ and the average rate of change would be $\Delta y/\Delta x=.64/.8=.8$. If the change in x were only $\Delta x=.5$, the change in y would be $\Delta y=0.55$, and the average rate of change would be $\Delta y/\Delta x=.55/.5=1.1$. As the change in x approaches 0 (i.e., $\Delta x\to0$), the rate of change $\Delta y/\Delta x$ approaches $\mathrm{d}y/\mathrm{d}x=1.6$. Thus, when $x_{\mathrm{i}}=.2$, $\mathrm{d}y/\mathrm{d}x=1.6$,

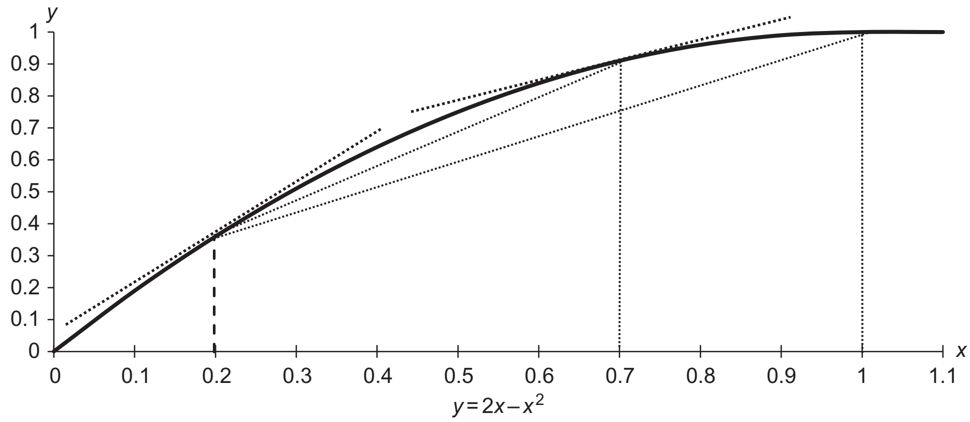

FIGURE 1.1 **The derivative of $y = 2x - x^2$.** When $x_i = .2$, $dy/dx = 1.6$. As $\Delta x \to 0$, $\Delta y/\Delta x \to dy/dx$. Also, notice that when $x_i = .7$, $dy/dx = .6$.

and an infinitesimal change in x would lead to 1.6 times that rate of change in y. The "point slope" or instantaneous rate of change of $2x - x^2$ is 1.6 when $x_i = .2$. The derivative of y with respect to $x(dy/dx = f'(x))$ can be interpreted to be the instantaneous rate of change in y given an infinitesimal change in x. In addition, notice that the slope (derivative) in Figure 1.1 changes with x. For example, when $x_i = .7$, $dy/dx = .6$. The rate of change of this derivative is the derivative of the derivative function, or the second derivative of the function $f(x)$ is $\frac{d^2y}{dx^2} = f''(x)$. In our example, $f''(x) = -2$. This means that the slope of the tangent line itself is changing at a constant rate of -2. Thus, after each change in x by 1 unit, the value of the slope will decrease by 2 units. This is apparent in Figure 1.1, since as x increases, the slope of the curve decreases.

1.4.1 Essential Rules for Calculating Derivatives

Equation (1.1) provides for a change in y given a very small (infinitesimal) change in x. This definition can be used to derive a number of very useful rules in calculus. A few are discussed below.

1.4.1.1 The Power Rule

One type of function that appears regularly in finance is the polynomial or integer power function. This type of function defines variable y in terms of a coefficient c, variable x, and an exponent n. While the exponents in a polynomial equation are non-negative integers, the rules that we discuss here still apply when the exponents assume negative or non-integer values. Consider a polynomial with a single variable x, a coefficient c, and an exponent n:

$$y = cx^n$$

The derivative of such a function y with respect to x is given by:

$$\frac{dy}{dx} = cnx^{n-1} \tag{1.2}$$

1.4.1.2 *The Sum Rule*

Consider a function that defines variable y in terms of a series of terms or functions involving x:

$$\frac{\mathrm{d}}{\mathrm{d}x}[f(x) + g(x)] = \frac{\mathrm{d}}{\mathrm{d}x}[f(x)] + \frac{\mathrm{d}}{\mathrm{d}x}[g(x)] \tag{1.3}$$

The notation $\frac{\mathrm{d}}{\mathrm{d}x}[f(x)]$ refers to the derivative of the function $f(x)$. In addition, the sum rule applies to any finite sum of terms. For example, consider y as a function of a series of coefficients c_j, variable x, and a series of exponents n_j:

$$y = \sum_{j=1}^{m} c_j \cdot x^{n_j} \tag{1.4}$$

The derivative of such a function y with respect to x is given by:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \sum_{j=1}^{m} c_j \cdot n_j \cdot x^{n_j - 1} \tag{1.5}$$

That is, simply take the derivative of each term in y with respect to x and sum these derivatives.

1.4.1.3 *The Chain Rule*

Each of the functions discussed in the previous section is written in polynomial form. Other rules can be derived to find derivatives for different types of functions. The *chain rule* is a derivative rule that allows us to differentiate more complex functions of the form:

$$y = f(g(x))$$

where *f(x)* and *g(x)* are functions whose derivatives are already known. The chain rule states that:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f'(g(x))g'(x) \tag{1.6}$$

To appreciate when the chain rule is relevant, consider the following two examples. First, consider $y = x^{1/2}$. We obtain the derivative as follows:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{2}x^{-1/2}$$

Next, consider the more complicated function $y = (x^3 + 4x - 1)^{1/2}$. We need to use the chain rule to find the derivative of y with respect to x. Observe that if we choose $f(x) = x^{1/2}$ and $g(x) = x^3 + 4x - 1$, then:

$$y = f(g(x)) = (g(x))^{1/2} = (x^3 + 4x - 1)^{1/2}$$

We already know how to find the derivatives: $f'(x) = \frac{1}{2}x^{-\frac{1}{2}}$ and $g'(x) = 3x^2 + 4$. Application of the chain rule to the composite function yields:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f'(g(x))g'(x) = \frac{1}{2}\left(x^3 + 4x - 1\right)^{-\frac{1}{2}}\left(3x^2 + 4\right)$$

Another way to express the chain rule is to create an intermediate variable, say u, with $u = g(x)$. If $y = f(g(x))$, then $y = f(u)$. With this notation, the chain rule can be expressed as:

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

Consider again the example $y = (x^3 + 4x - 1)^{1/2}$. Choose $u = x^3 + 4x - 1$, so that $y = u^{1/2}$. By using the chain rule, we obtain:

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = \frac{1}{2}u^{-1/2}\left(3x^2 + 4\right) = \frac{1}{2}(x^3 + 4x - 1)^{-1/2}\left(3x^2 + 4\right)$$

Consider one more example where $y = x^3$ and $x = t^2 + 1$ and we wish to find dy/dt. Again, from the chain rule, we have:

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt} = 3x^2(2t) = 3\left(t^2+1\right)^2(2t) = 6t\left(t^2 + 1\right)^2$$

1.4.1.4 Product and Quotient Rules

The *product rule*, which is applied to a function such as $y = f(x)g(x)$, holds that the derivative of y with respect to x is as follows:

$$\frac{dy}{dx} = f(x)\frac{dg(x)}{dx} + g(x)\frac{df(x)}{dx} \tag{1.7}$$

For example, if $y = (4x + 2)(5x + 1)$ where $f(x)$ is $(4x + 2)$ and $g(x)$ is $(5x + 1)$, the product rule holds that $dy/dx = (4x + 2) \times 5 + (5x + 1) \times 4 = 40x + 14$.

The *quotient rule*, which is applied to a function such as $f(x)/g(x)$, holds that the derivative of y with respect to x is as follows:

$$\frac{dy}{dx} = \left[g(x)\frac{df(x)}{dx} - f(x)\frac{dg(x)}{dx}\right]/g(x)^2 \tag{1.8}$$

For example, if $y = (4x + 2)/5x$ where $f(x)$ is $(4x + 2)$ and $g(x)$ is $5x$, the quotient rule holds that $dy/dx = [(5x \times 4) - 5(4x + 2)]/25x^2 = -2/5x^2$.

The product rule also implies the constant multiple rule:

$$\frac{d}{dx}\left[cf(x)\right] = c\frac{d}{dx}\left[f(x)\right] \quad \text{(constant multiple rule)} \tag{1.9}$$

1.4.1.5 Exponential and Log Function Rules

Logarithmic and exponential functions and derivatives of these functions are particularly useful in finance for modeling growth. Consider the function $y = e^x$ and its derivative with respect to x:

$$\frac{dy}{dx} = e^x \tag{1.10}$$

Or, more generally, which can be verified with the chain rule:

$$\frac{de^{g(x)}}{dx} = \frac{dg(x)}{dx}e^{g(x)} \tag{1.11}$$

If $y = e^{\ln(x)}$, then, by definition, $y = e^{\ln(x)} = x$, which implies that $de^{\ln(x)}/dx = 1$. Now, consider the following special case of Eq. (1.11):

$$\frac{de^{\ln(x)}}{dx} = \frac{d\ln(x)}{dx} e^{\ln(x)}$$

which implies:

$$\begin{aligned} 1 &= \frac{d\ln(x)}{dx} \cdot x \\ \frac{d\ln(x)}{dx} &= \frac{1}{x} \end{aligned} \tag{1.12}$$

Table 1.1 summarizes the rules for finding derivatives covered in Section 1.4.1. We will make regular use of these rules throughout the text.

1.4.2 The Differential

The concept of the differential will be very useful later when we discuss stochastic calculus. The differential of a function can be used to estimate the change of the value of a function $y = f(x)$ resulting from a small change of the x value. Since:

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

then when Δx is small we have:

$$f'(x) \cong \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

The approximation improves as Δx approaches 0. Denote the error in the approximation above by $\epsilon(x, \Delta x)$, so that:

$$\frac{f(x + \Delta x) - f(x)}{\Delta x} = f'(x) + \epsilon(x, \Delta x)$$

TABLE 1.1 Sample Derivative Rules (c and n are Arbitrary Constrants)

1. $\frac{d}{dx}[x^n] = nx^{n-1}$ (power rule)	**5.** $\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{g(x)\frac{d}{dx}[f(x)] - f(x)\frac{d}{dx}[g(x)]}{[g(x)]^2}$ (quotient rule)
2. $\frac{d}{dx}[f(x) + g(x)] = \frac{d}{dx}[f(x)] + \frac{d}{dx}[g(x)]$ (sum rule)	**6.** $\frac{d}{dx}\left[cf(x) = c\frac{d}{dx}[f(x)]\right.$ (constant multiple rule)
3. $\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$ (chain rule)	**7.** $\frac{d}{dx}[e^x] = e^x$ (exponential rule)
4. $\frac{d}{dx}[f(x)g(x)] = f(x)\frac{d}{dx}[g(x)] + g(x)\frac{d}{dx}[f(x)]$ (product rule)	**8.** $\frac{d}{dx}[\ln x] = \frac{1}{x}$ (log rule)

Whenever the derivative $f'(x)$ exists, this equality and our definition above for $f'(x)$ imply that $\epsilon(x,\Delta x) \rightarrow 0$ as $\Delta x \rightarrow 0$. Now, we label the change in y by Δy, so that:

$$\Delta y = f(x + \Delta x) - f(x) = f'(x)\Delta x + \epsilon(x,\Delta x)\Delta x$$

Observe on Figure 1.2 that Δy, the change in y on the curve, can be closely approximated by $f'(x)\Delta x$ when Δx is small. The expression $f'(x)\Delta x$ is the change in y on the tangent line resulting from the change Δx in the value of x. In the case that $\epsilon(x,\Delta x) \rightarrow 0$ as $\Delta x \rightarrow 0$ (so that the error term is negligible as Δx approaches 0), then one often writes:

$$\mathrm{d}y = f'(x)\mathrm{d}x$$

where $\mathrm{d}x$ has replaced Δx and $\mathrm{d}y$ has replaced Δy. The term $\mathrm{d}y$ is called the differential of y.

ILLUSTRATION: THE DIFFERENTIAL AND THE ERROR

Reconsider our illustration from earlier with $y = 2x - x^2$, plotted again in Figure 1.2. The differential $\mathrm{d}y = (2 - 2x)\mathrm{d}x$. Suppose that in this case $x = .6$ and $\mathrm{d}x = 0.1$, such that $\mathrm{d}y = (2 - 2 \times .6)(.1) = .08$. This tells us that the approximate change in y from $x = .6$ by $\Delta x = .1$ to $x = .7$ will be $\Delta y \approx .08$. The actual change in y can be computed directly since:

$$f(.7) - f(.6) = \left(1.4 - .7^2\right) - \left(1.2 - .6^2\right) = .07$$

The term $\epsilon(x,\Delta x)\Delta x$ itself is the error in using the differential as an approximation to the change in y. More precisely:

$$\epsilon(x,\Delta x)\Delta x = \left[f(x + \Delta x) - f(x)\right] - f'(x)\Delta x$$

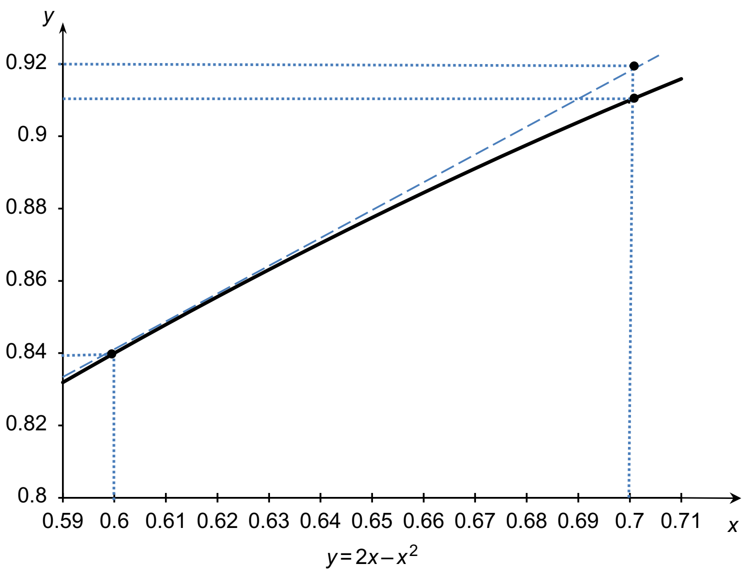

FIGURE 1.2 **The differential.** When $x_0 = .6$, $\mathrm{d}y/\mathrm{d}x = .8$. As $\Delta x \rightarrow 0$, $\Delta y/\Delta x \rightarrow \mathrm{d}y/\mathrm{d}x$. Also, where the tangent dashed line reflects error estimates for y based on error estimates for x and $\mathrm{d}y$, notice that when $x = .7$, $\epsilon\Delta x = .01$.

where Δx is regarded as the same as dx. In our example we have:

$$\epsilon(.6,.1)\Delta x = \left[f(.7) - f(.6)\right] - f'(.6) \times .1 = [.91 - .84] - .08 = .07 - .08 = -.01$$

Observe that the possible error $\epsilon(.6,.1)\Delta x = -.01$ is very small relative to the differential d$y = .08$. Observe that the differential provided a reasonable estimate, and the error term $\epsilon(.6, .1)\Delta x$, because of its size relative to Δx, can be ignored as Δx approaches 0.

1.4.3 Partial Derivatives

If our dependent variable y is a function of multiple independent variables x_j, we can find partial derivatives $\partial y/\partial x_j$ of y with respect to each of our independent variables x_j. For example, in the following, y is a function of x_1 and x_2; function y's partial derivatives with respect to each of its independent variables (while holding the other constant) follow:

$$y = x_1 e^{.05x_2} + .03x_2$$
$$\frac{\partial y}{\partial x_1} = e^{.05x_2}$$
$$\frac{\partial y}{\partial x_2} = .05x_1 e^{.05x_2} + .03$$

1.4.3.1 The Chain Rule for Two Independent Variables

Suppose that $y = f(x)$ and $x = g(t)$. Recall that the chain rule provides:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{\mathrm{d}y}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}t}$$

There is an analogous chain rule for functions of more than one independent variable. Suppose the variable z is a function of the variables x and y, $z = f(x, y)$, and, in turn, each of the variables x and y is a function of the variable t, $x = g(t)$ and $y = h(t)$. This implies that z can be defined as function of the variable t, that is $z = f(g(t), h(t))$.

Now, consider an example where $z = x^2y + y^3$, $x = t^4$, and $y = 2t$. This implies that $z = (t^4)^2(2t) + (2t)^3 = 2t^9 + 8t^3$. While the derivative $\frac{\mathrm{d}z}{\mathrm{d}t} = 18t^8 + 24t^2$ can easily be obtained by a direct calculation from this last expression, it can also be found using the chain rule. Since $z = f(x, y)$, $x = g(t)$, and $y = h(t)$, the derivative $\frac{\mathrm{d}z}{\mathrm{d}t}$ is obtained from the chain rule:

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \frac{\partial z}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial z}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} = 2xy\left(4t^3\right) + (x^2 + 3y^2)(2)$$
$$= 2t^4(2t)\left(4t^3\right) + \left[\left(t^4\right)^2 + 3(2t)^2\right](2) = 18t^8 + 24t^2$$

Observe that we obtained the same answer earlier by the direct calculation.

From the chain rule, we multiply through by dt to derive the total differential:

$$\mathrm{d}z = \frac{\partial z}{\partial x}\mathrm{d}x + \frac{\partial z}{\partial y}\mathrm{d}y$$

We will demonstrate in later chapters that the total differential is a useful tool to find solutions to certain types of differential equations. It is also useful for approximating the change in the variable z (dz) resulting from small changes in the variables x and y (dx and dy).

1.4.4 Taylor Polynomials and Expansions

One can improve on the approximation $\Delta y = f'(x)\Delta x$ by taking into account higher-order derivatives. As long as the function $f(x)$ is differentiable at least n times, the nth-order Taylor polynomial expanded about x_0 is defined by the right side of the following approximation:

$$f(x_0 + \Delta x) \approx f(x_0) + f'(x_0)(\Delta x) + \frac{1}{2!}f''(x_0)(\Delta x)^2 + \frac{1}{3!}f^{(3)}(x_0)(\Delta x)^3 + \cdots + \frac{1}{n!}f^{(n)}(x_0)(\Delta x)^n$$

The Taylor polynomial approximation or expansion can be used for finite changes in x. Taylor polynomial expansions are frequently used to evaluate a function $f(x)$ at some point x_1 that differs from an initial point x_0 at which $f(x_0)$ has already been evaluated. That is, the Taylor polynomial can be used to approximate a change in $f(x)$ induced by a change in x. For example, consider the function $f(x) = \ln x$. Choose $x_0 = 1$. We will use the third-order Taylor polynomial to estimate the value of $\ln(1.2)$. Differentiating, we obtain $f'(x) = x^{-1}$, $f''(x) = -x^{-2}$, and $f^{(3)}(x) = 2x^{-3}$.[6] Evaluating at $x_0 = 1$ yields $f(1) = 0$, $f'(1) = 1$, $f''(1) = -1$, and $f^{(3)}(1) = 2$. In this case, $x_1 = 1.2$, such that $\Delta x = 1.2 - 1 = .2$. We obtain our estimate for $f(x_0 + \Delta x)$ as follows:

$$f(x_0 + \Delta x) = \ln(1 + .2) = \ln(1.2) \approx 0 + 1(.2) + \frac{1}{2}(-1)(.2)^2 + \frac{1}{6}(2)(.2)^3 = .182667$$

The actual value of $\ln(1.2)$ is .18232. ... Observe that the first-order Taylor polynomial would yield the estimate .2, and the second-order would give .18. In general, the higher the order of the Taylor polynomial, the better the estimate. We are often concerned with changes in the value of the function $f(x)$. Since $\Delta y = f(x + \Delta x) - f(x)$, then after replacing x_0 with x and subtracting $f(x)$, we can express the previous approximation as:

$$\Delta y = f'(x)(\Delta x) + \frac{1}{2!}f''(x)(\Delta x)^2 + \frac{1}{3!}f^{(3)}(x)(\Delta x)^3 + \cdots + \frac{1}{n!}f^{(n)}(x)(\Delta x)^n$$

One can generalize the results above to functions of more than one independent variable. Consider the function $y = f(x,t)$. Define $\Delta y = f(x + \Delta x, t + \Delta t)) - f(x,t)$. This can also be expanded into a two-variable Taylor series where the first few terms out to the second-order derivatives in the expansion take the following form:

$$\Delta y = \frac{\partial f}{\partial x}(x,t)\Delta x + \frac{\partial f}{\partial t}(x,t)\Delta t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x,t)(\Delta x)^2 + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}(x,t)(\Delta t)^2 + \frac{\partial^2 f}{\partial x \partial t}(x,t)\Delta x \Delta t + \cdots$$

For example, consider the function $f(x,t) = 100e^{-x^2+3t}$. In this illustration, we will use a second-order Taylor polynomial to estimate the change $f(.2,.1) - f(0,0)$. So, we must choose $x = 0$, $t = 0$, $\Delta x = .2$ and $\Delta t = .1$. We find first and second derivatives as follows:

$$\frac{\partial f}{\partial x} = -200x\mathrm{e}^{-x^2+3t}, \quad \frac{\partial f}{\partial t} = 300\mathrm{e}^{-x^2+3t}$$

$$\frac{\partial^2 f}{\partial x^2} = 200(-1 + 2x^2)\mathrm{e}^{-x^2+3t}, \quad \frac{\partial^2 f}{\partial x \partial t} = -600x\mathrm{e}^{-x^2+3t}, \quad \frac{\partial^2 f}{\partial t^2} = 900\mathrm{e}^{-x^2+3t}$$

Evaluating the derivatives at $(x,t) = (0,0)$:

$$\frac{\partial f}{\partial x}(0,0) = 0, \ \frac{\partial f}{\partial t}(0,0) = 300, \ \frac{\partial^2 f}{\partial x^2}(0,0) = -200, \frac{\partial^2 f}{\partial x \partial t}(0,0) = 0, \ \frac{\partial^2 f}{\partial t^2}(0,0) = 900$$

Thus,

$$\Delta f = f(.2,.1) - f(0,0) \cong 0(.2) + 300(.1) + \frac{1}{2}(-200)(.2)^2 + \frac{1}{2}(900)(.1)^2 + (0)(.2)(.1)$$
$$= 30.5$$

1.4.5 Optimization and the Method of Lagrange Multipliers

Differential calculus is particularly useful for determining minima or maxima of functions of many types. In many instances, minima or maxima can be calculated by setting first derivatives with respect to the variable(s) of interest equal to zero (first-order conditions), and then checking second-order conditions (positive second derivative(s) for minima, negative second derivative(s) for maxima).

However, many optimization problems require constraints or limitations on variables. The method of Lagrange multipliers can often enable function optimization in the presence of such constraints. The method of Lagrange multipliers creates a Lagrange function L that supplements the original function $y = f(x)$ to be optimized with an additional expression for each of m relevant constraints. Assume a linear constraint equation of the form $\mathbf{g}(\mathbf{x}) = \mathbf{c}$, with $\mathbf{g}$ an $m \times 1$ vector valued function of the vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$, which is an $n \times 1$ column vector variable, and $\mathbf{c}$ is an $m \times 1$ constant column vector. We will introduce the Lagrange multiplier column vector $\boldsymbol{\lambda}$ where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)^{\mathrm{T}}$. The Lagrange function has the form:

$$L = f(\mathbf{x}) + \boldsymbol{\lambda}^{\mathrm{T}}(\mathbf{g}(\mathbf{x}) - \mathbf{c})$$

ILLUSTRATION: LAGRANGE OPTIMIZATION

Suppose that our objective is to minimize the function $y = x_1{}^2 + 2x_2^2 + .5x_1x_2$ subject to the constraint that $x_1 + .2x_2 = 10$:[7]

$$\text{OBJ : Min } y = x_1^2 + 2x_2^2 + .5x_1x_2$$
$$\text{s.t. : } x_1 + .2x_2 = 10$$

The Lagrange function combines the original function and a revised version of the single constraint as follows:[8]

$$L = x_1^2 + 2x_2^2 + .5x_1x_2 + \lambda(x_1 + .2x_2 - 10)$$

We solve our problem by setting partial derivatives of L with respect to each of our three variables equal to zero. This will result in the following first-order conditions:

$$\frac{\partial L}{\partial x_1} = 2x_1 + .5x_2 + \lambda = 0$$

$$\frac{\partial L}{\partial x_2} = .5x_1 + 4x_2 + .2\lambda = 0$$

$$\frac{\partial L}{\partial \lambda} = x_1 + .2x_2 - 10 = 0$$

This system is structured and solved in matrix format as follows:

$$\begin{bmatrix} 2 & .5 & 1 \\ .5 & 4 & .2 \\ 1 & .2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 10 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} .010309 & -.05155 & 1.005155 \\ -.05155 & .257732 & -.02577 \\ 1.005155 & -.02577 & -1.99742 \end{bmatrix} \times \begin{bmatrix} 0 \\ 0 \\ 10 \end{bmatrix} = \begin{bmatrix} 10.05155 \\ -.25773 \\ -19.9742 \end{bmatrix}$$

Thus, y is minimized when $x_1 = 10.05155$ and $x_2 = -.25773$. The Lagrange multiplier λ can be interpreted as a sensitivity coefficient that indicates the change in y that would result from a change in the constraint on $x_1 + .2x_2$. If, for example, we were to increase the constraint by 1 from 10 to 11, the value of y would decrease by approximately 19.9742.

1.5 REVIEW OF INTEGRAL CALCULUS

A graphic interpretation of the derivative $f'(x)$ of a function $f(x)$ is that it equals the slope of the curve plotted by that function. A graphic interpretation of the integral of a non-negative function $f(x)$, $\int_a^b f(x)\mathrm{d}x$, is that it equals the area under the graph of the function $f(x)$ from $x = a$ to $x = b$, where $\int$ is the *integral sign* and $f(x)$ is the *integrand*. Thus, integrals are useful for finding areas under curves. Integrals can be regarded as the limit of sums involving functions of a continuous variable. Similarly, as we will discuss shortly, they are useful for determining expected values and variances based on continuous probability distributions. As the expectation of a discrete random variable requires summing a discrete countable number of terms, the expectation of a continuous random variable requires integration to handle the continuous (uncountable) number of values of the random variable.

Integral calculus is also useful for analyzing the behavior of variables (such as cash flows) over time. An equation of the form $\frac{\mathrm{d}y}{\mathrm{d}t} = f(t)$ is known as a differential equation and it might describe the rate of change of the variable y with respect to time t. The solution to this differential equation $y = F(t)$, which is obtained by integration, describes the function y itself over time. For example, $f(t)$ might describe the change in value of the price y of an investment over time (profit) while $F(t)$ provides the actual value of the price.

1.5.1 Antiderivatives

Integrals of many functions can be determined by using the process of *antidifferentiation*, which is the inverse process of differentiation. If $F(x)$ is a function of x whose derivative equals $f(x)$, then $F(x)$ is said to be the antiderivative or integral of $f(x)$, written as follows:

$$F(x) = \int f(x)\,\mathrm{d}x \tag{1.13}$$

The function $F(x)$ has the property that:

$$\frac{\mathrm{d}F(x)}{\mathrm{d}x} = f(x) \tag{1.14}$$

One can always add any constant C to the function $F(x)$, where $F(x)$ is any one particular antiderivative of $f(x)$, and it will still be an antiderivative of $f(x)$; that is:

$$\frac{\mathrm{d}}{\mathrm{d}x}[F(x) + C] = \frac{\mathrm{d}}{\mathrm{d}x}[F(x)] + \frac{\mathrm{d}}{\mathrm{d}x}[C] = f(x) + 0 = f(x)$$

Thus, the general form of the *indefinite integral* of $f(x)$ is:

$$\int f(x)\mathrm{d}x = F(x) + C$$

where $F(x)$ is one particular antiderivative of $f(x)$. Observe that the indefinite integral of a function is actually a family of functions, since each different choice of the constant C gives a different function.

Suppose, for example, we wished to evaluate $\int 2x\,\mathrm{d}x$. We will seek a family of functions for which the derivative is $2x$. Since $\frac{\mathrm{d}}{\mathrm{d}x}\left[x^2 + C\right] = 2x$, $\int 2x\,\mathrm{d}x = x^2 + C$. Using the fact that integrals are the inverse process of differentiation, one can derive integral rules. Table 1.2 provides a short listing of integral rules that will be useful in this book.

TABLE 1.2 Table of Integrals

1.	$\int cx^n\,\mathrm{d}x = \frac{cx^{n+1}}{n+1} + C$ for $n \neq -1$ (power rule)
2.	$\int cf(x)\mathrm{d}x = c\int f(x)\mathrm{d}x$ (constant multiple rule)
3.	$\int (f(x) + g(x))\mathrm{d}x = \int f(x)\mathrm{d}x + \int g(x)\mathrm{d}x$ (sum rule)
4.	$\int \frac{1}{x}\mathrm{d}x = \ln\lvert x\rvert + C$
5.	$\int \mathrm{e}^{cx}\,\mathrm{d}x = \frac{1}{c}\mathrm{e}^{cx} + C$

Next, suppose that we wished to evaluate $\int\left(\frac{5}{x}+3e^x+4x^2-6\right)dx$. We will use all five rules in Table 1.2 to evaluate this function, finding that $\int\left(\frac{5}{x}+3e^x+4x^2-6\right)dx = 5\ln|x| + 3e^x + \frac{4}{3}x^3 - 6x + C$. Observe that there is only one constant C in the solution. This is sufficient since C can be any arbitrary constant.

1.5.2 Definite Integrals

Using simple rules from geometry, one can find areas of elementary shapes such as squares, rectangles, triangles, and circles. However, if we wish to find the area under the graph of an arbitrary curve, we need a new method. If the values a function $f(x)$ are non-negative so that its graph always lies above the x-axis, then the definite integral of $f(x)$ from $x=a$ to $x=b$ is defined to be the area between the x-axis and its graph from $x=a$ to $x=b$ (see Figure 1.3). For a general function $f(x)$, the definite integral from $x=a$ to $x=b$ equals the area above the x-axis minus the area below the x-axis. The definite integral is denoted by:

$$\int_a^b f(x)\,dx$$

Notice that the notation for this area uses the antiderivative sign. We will show this connection shortly by using the fundamental theorem of calculus.

1.5.2.1 Reimann Sums

The definite integral for any continuous curve can be obtained as a limit of a sum of rectangular areas (or so-called "signed areas" in case that part of the graph of $f(x)$ is below the x-axis). More precisely, consider the graph of a function $f(x)$ on the interval $[a,b]$ of x-values. For the time being, suppose that the function $f(x)$ is non-negative. Divide the interval $[a,b]$ into n subintervals of equal width

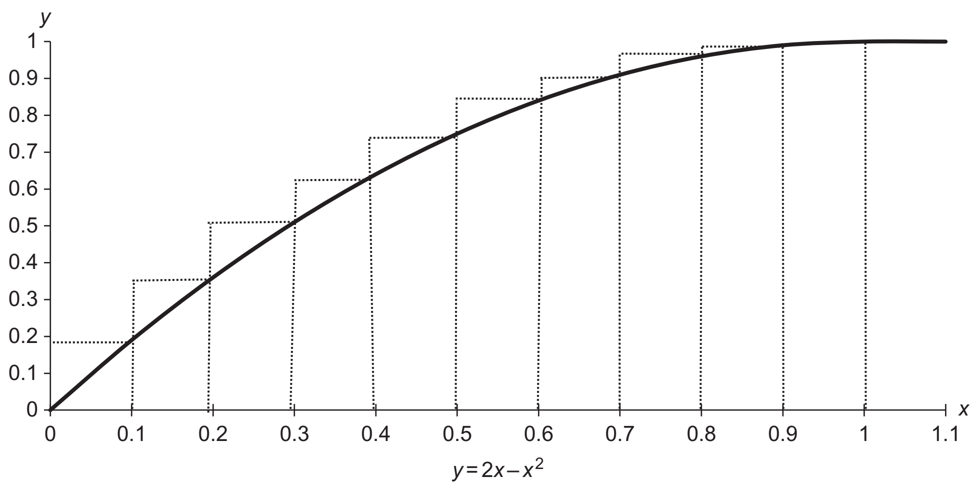

FIGURE 1.3 **The area under $f(x) = 2x - x^2$.** When $x_i - x_{i-1} = .1$, the sum of the areas of the 10 rectangles equals 0.715. As the number of rectangles approaches infinity, and their widths approach zero, the sum of their areas will approach 2/3.

$\Delta x = (b - a)/n$. Consider the values of x on the x-axis that are endpoints of the subintervals. They are: $x_0 = a$, $x_1 = a + \Delta x$, $x_2 = a + 2\Delta x$, ..., $x_n = a + n\Delta x = b$. The n subintervals are: $[x_0,x_1]$, $[x_1,x_2]$, ..., $[x_{n-1},x_n]$. For the ith subinterval $[x_{i-1},x_i]$, choose a convenient x-value x_i* in this subinterval; that is, $x_{i-1} \le x_i^* \le x_i$. The area between the graph of $f(x)$ and the x-axis can be approximated by the sum of the areas of the n rectangles so that the ith rectangle has height $f(x_i^*)$ and width Δx. Since the area of a rectangle is the product of its height and width, the area of the ith rectangle equals $f(x_i^*)\Delta x$. Thus, the total area from $x = a$ to $x = b$ can be approximated by the sum:

$$\sum_{i=1}^{n} f\left(x_i^*\right)\Delta x$$

This sum is known as a *Riemann sum*. To illustrate this Riemann sum, consider the example of estimating the area under the graph of $y = 2x - x^2$ from $x = 0$ to $x = 1$. We choose $n = 10$ so that we are estimating the area under the graph by 10 narrow rectangles (see Figure 1.3). In this case the width of each rectangle is $\Delta x = (1 - 0)/10 = .1$. The value of $x_i = i/10$. For this example, choose $x_i^* = x_i$. The sum of the areas of the 10 narrow rectangles equals the following Riemann sum, as calculated in Table 1.3:

$$\sum_{i=1}^{10}(2x_i - x_i^2)\Delta x = \sum_{i=1}^{10}\left[\frac{2i}{10} - \left(\frac{i}{10}\right)^2\right].1 = .715$$

Now suppose that we increase the number of rectangles from $n = 10$ an arbitrarily large value of n. For general n, the width of the each rectangle $\Delta x = 1/n$. The ith x-value is $x_i = i/n$. For simplicity, choose $x_i^* = x_i = i/n$. Clearly as n approaches infinity, the Riemann sum should approach

TABLE 1.3 Riemann Sums and Calculating the Area Under $f(x) = 2x - x^2$

i	$f(x_i)$	x_i	Δx	$f(x_i)\Delta x$
1	0.19	.1	0.1	0.019
2	0.36	.2	0.1	0.036
3	0.51	.3	0.1	0.051
4	0.64	.4	0.1	0.064
5	0.75	.5	0.1	0.075
6	0.84	.6	0.1	0.084
7	0.91	.7	0.1	0.091
8	0.96	.8	0.1	0.096
9	0.99	.9	0.1	0.099
10	1	1	0.1	0.100
			$\sum f(x_i)\Delta x =$	**0.715**

the exact area between the curve $y = 2x - x^2$ and the x-axis from $x = 0$ to $x = 1$. So the exact area will equal:

$$\lim_{n\to\infty}\sum_{i=1}^{n} f(x_i)\Delta x = \lim_{n\to\infty}\sum_{i=1}^{n}\left[\frac{2i}{n} - \left(\frac{i}{n}\right)^2\right]\frac{1}{n}$$

Some algebraic manipulation reveals that the sum on the right equals $\frac{2}{3} + \frac{1}{2n} - \frac{1}{6n^2}$. Thus, the exact area under the curve is 2/3 since $\frac{1}{2n} - \frac{1}{6n^2}$ approaches 0 as n approaches infinity.

Recall that this area is defined to be the definite integral of $f(x) = 2x - x^2$ from $x = 0$ to $x = 1$. This can be expressed as:

$$\text{Area} = \int_0^1 \left(2x - x^2\right)\,\mathrm{d}x = \frac{2}{3}$$

In general, for any non-negative continuous function $f(x)$, we can express the area as:

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{n\to\infty}\sum_{i=1}^{n} f\left(x_i^*\right)\Delta x$$

Thus, as the number of subintervals Δx increases, and the widths of each narrow, the area under $f(x)$ approaches the limit of the sum of the rectangular areas. Note that if the graph of $f(x)$ extends below the x-axis, where the values of $f(x)$ are negative, then the terms $f(x_i^*)\Delta x$ are negative, and the terms $f(x_i)\Delta x$ are negative in equation x, such that the contribution to the definite integral will be negative for this portion of the graph. Thus, in general, the definite integral of any continuous function $f(x)$ equals the area of the region above the x-axis minus the area of the region below the x-axis.

It can be challenging or even impossible to compute the right-hand sum of x_i^* and find its limit. The powerful fundamental theorem of calculus often allows us to easily find these areas for a wide range of functions:

Fundamental theorem of calculus: If f(x) is any continuous function on the interval [a,b], then

$$\int_a^b f(x)\,\mathrm{d}x = F(x)\Big|_a^b = F(b) - F(a)$$

where F(x) is any particular antiderivative of f(x).

The essential steps of the proof of this theorem can be found in the companion website.

Recall our earlier example, find the area under the graph of $y = 2x - x^2$ from $x = 0$ to $x = 1$. Using the fundamental theorem of calculus, the area equals:

$$\int_0^1 (2x - x^2)\,\mathrm{d}x = x^2 - \frac{1}{3}x^3\Big|_0^1 = \left[1^2 - \frac{1}{3}1^3\right] - \left[0^2 - \frac{1}{3}0^3\right] = \frac{2}{3}$$

The notation $F(x)|_a^b$ is equivalent to $F(b) - F(a)$, but it is useful to use this notation as an intermediate step in evaluating definite integrals, since one first finds the antiderivative $F(x)$ before evaluating $F(x)$ at $x = b$ and $x = a$ and then finally taking their difference. Observe that this is what we did in the example above.

We note that the definite integral is independent of the particular antiderivative that is chosen. To illustrate this point, we recalculate the above definite integral allowing for different choices of the antiderivative:

$$\int_0^1 (2x - x^2)\,\mathrm{d}x = x^2 - \frac{1}{3}x^3 + C\Big|_0^1 = \left[1^2 - \frac{1}{3}1^3 + C\right] - \left[0^2 - \frac{1}{3}0^3 + C\right] = \frac{2}{3}$$

As we shall discuss in Section 4.1.3, the importance of Riemann sums and their limits extends beyond their applications to finding areas under curves. Many continuous valuation models are based on Riemann sums and their limits as the widths of the horizontal intervals approach zero.

1.5.3 Change of Variables Technique to Evaluate Integrals

An important technique for evaluating integrals is a change of variables. This substitution technique can significantly reduce the apparent complexity of many functions. Suppose we wish to integrate some function of the variable t. Suppose that we can choose a new variable x that is a function of t ($x = x(t)$) in just the right way, so that the integral takes the form:

$$\int f(x(t))\frac{\mathrm{d}x}{\mathrm{d}t}\,\mathrm{d}t$$

In this case, one can change the variable of integration from the variable t to the variable x and integrate the function $f(x)$ to evaluate the integral. Once the integral has been evaluated, one simply substitutes in place of the variable x the function $x(t)$. To express this symbolically:

$$\int f(x(t))\frac{\mathrm{d}x}{\mathrm{d}t}\,\mathrm{d}t = \int f(x)\,\mathrm{d}x = F(x) + C = F(x(t)) + C$$

where $F(x)$ is an antiderivative of $f(x)$.

The proof of this result is quite simple. Suppose $\int f(x)\mathrm{d}x = F(x) + C$ is the general antiderivative of $f(x)$. If we differentiate $F(x(t)) + C$ in the variable t, by the chain rule we have:

$$\frac{\mathrm{d}}{\mathrm{d}t}[F(x(t)) + C] = \frac{\mathrm{d}}{\mathrm{d}x}[F(x) + C]\frac{\mathrm{d}x}{\mathrm{d}t} = f(x)\frac{\mathrm{d}x}{\mathrm{d}t} = f(x(t))\frac{\mathrm{d}x}{\mathrm{d}t}$$

Thus, we have proved that the derivative of $F(x(t)) + C$ in the variable t equals $f(x(t))\frac{\mathrm{d}x}{\mathrm{d}t}$, thus the antiderivative of $f(x(t))\frac{\mathrm{d}x}{\mathrm{d}t}$ in the variable t equals $F(x(t)) + C$, as we wanted to prove.

In order to make use of the technique of change of variables, one needs to be able to find the right choice for the function x in terms of t. This is a matter of practice and being able to visualize the function $x(t)$ in the expression one is attempting to integrate. We should say that the change of variables method only works if the function one is integrating is able to be expressed in the special form $f(x(t))\frac{\mathrm{d}x}{\mathrm{d}t}$ and one can find the antiderivative of $f(x)$ in the variable x. We also point out that in calculus textbooks the change of variables method is often called u-substitution. This is because in calculus textbooks the substitution variable is often denoted by u.

ILLUSTRATION: CHANGE OF VARIABLES TECHNIQUE FOR THE INDEFINITE INTEGRAL

Suppose that we seek to evaluate the indefinite integral $\int (t^2+1)^3 t\,\mathrm{d}t$. First, we notice that the quantity t immediately to the left of dt is almost the derivative of t^2+1, which is the base of the cubed function in the parentheses. This motivates the attempt to choose $x = t^2 + 1$, such that $\frac{\mathrm{d}x}{\mathrm{d}t} = 2t$. We now do a little bit of algebra to create the expression that we need:

$$\int (t^2+1)^3 t\,\mathrm{d}t = \int \frac{1}{2}(t^2+1)^3(2t)\,\mathrm{d}t$$

So, in this case $f(x) = \frac{1}{2}x^3$ and $x = t^2 + 1$. Observe that $\frac{\mathrm{d}x}{\mathrm{d}t} = 2t$ such that $\mathrm{d}x = 2t\ \mathrm{d}t$. We rewrite as follows:

$$\int \frac{1}{2}(t^2+1)^3(2t)\,\mathrm{d}t = \int \frac{1}{2}x^3\,\mathrm{d}x = \frac{1}{8}x^4 + C = \frac{1}{8}(t^2+1)^4 + C$$

1.5.3.1 Change of Variables Technique for the Definite Integral

For definite integrals, the change of variables method is the same to determine the antiderivative. The only additional feature is that one can also express the limits of integration in terms of the new variable. If the limits of integration in the variable t are from a to b, then the limits of integration in the variable $x = x(t)$ will be from $x(a)$ to $x(b)$. Thus:

$$\int_a^b f(x(t))\frac{\mathrm{d}x}{\mathrm{d}t}\,\mathrm{d}t = \int_{x(a)}^{x(b)} f(x)\,\mathrm{d}x$$

Now, suppose that we seek to evaluate the indefinite integral $\int_0^1 (t^2+1)^3 t\,\mathrm{d}t$. We already calculated the integral of this function in the previous example, and now only have only left to evaluate the integral at the endpoints. Since $x = t^2 + 1$, then $x(0) = 1$ and $x(1) = 2$.

$$\int_0^1 (t^2+1)^3 t\,\mathrm{d}t = \frac{1}{8}x^4\bigg|_{x(0)}^{x(1)} = \frac{1}{8}x^4\bigg|_1^2 = \frac{2^4}{8} - \frac{1^4}{8} = \frac{15}{8}$$

We also could have obtained the solution by expressing the evaluated integral in terms of the original variable t, and then evaluating the integral at the endpoints in terms of t:

$$\int_0^1 (t^2+1)^3 t\,\mathrm{d}t = \frac{1}{8}(t^2+1)^4\bigg|_0^1 = \frac{1}{8}(1^2+1)^4 - \frac{1}{8}(0^2+1)^4$$
$$= \frac{15}{8}$$

1.6 EXERCISES

1.1. Add the following matrices:

$$\underset{\mathbf{A}}{\begin{bmatrix} 2 & 4 & 9 \\ 6 & 4 & 25 \\ 0 & 2 & 11 \end{bmatrix}} + \underset{\mathbf{B}}{\begin{bmatrix} 3 & 0 & 6 \\ 2 & 1 & 3 \\ 7 & 0 & 4 \end{bmatrix}} =$$

1.2. Subtract **E** from **D**:

$$\underset{\mathbf{D}}{\begin{bmatrix} 9 & 4 & 9 \\ 6 & 4 & 8 \\ 5 & 2 & 9 \end{bmatrix}} - \underset{\mathbf{E}}{\begin{bmatrix} 5 & 0 & 6 \\ 2 & 1 & 6 \\ 5 & 0 & 9 \end{bmatrix}} =$$

1.3. Transpose the following:

a. $$\underset{\mathbf{A}}{\begin{bmatrix} 1 & 8 & 9 \\ 6 & 4 & 25 \\ 3 & 2 & 35 \end{bmatrix}}$$

b. $$\underset{\mathbf{y}}{\begin{bmatrix} 9 \\ 6 \\ 3 \\ 7 \end{bmatrix}}$$

c. $$\underset{\mathbf{V}}{\begin{bmatrix} .09 & .01 & .04 \\ .01 & .16 & .10 \\ .04 & .10 & .64 \end{bmatrix}}$$

1.4. Multiply the following:

$$\underset{\mathbf{A}}{\begin{bmatrix} 7 & 4 & 9 \\ 6 & 4 & 12 \\ 3 & 2 & 17 \end{bmatrix}} \times \underset{\mathbf{B}}{\begin{bmatrix} 7 & 6 \\ 5 & 1 \\ 9 & 12 \end{bmatrix}}$$

1.5. Suppose that $\mathbf{A} = \begin{bmatrix} -2 & 0 \\ 3 & 4 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 7 & 3 \\ 5 & -1 \end{bmatrix}$. Find the following:

a. $2\mathbf{A}$
b. $\mathbf{A}^{\mathrm{T}}$
c. $\mathbf{A} + \mathbf{B}$
d. $\mathbf{AB}$
e. $\mathbf{BA}$

1.6. Invert the following matrices:

a. $\begin{bmatrix} 8 \end{bmatrix}$

b. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

c. $\begin{bmatrix} 4 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$

d. $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$

e. $\begin{bmatrix} .02 & .04 \\ .06 & .08 \end{bmatrix}$

f. $\begin{bmatrix} -2 & 1 \\ 1.5 & -.5 \end{bmatrix}$

g. $\begin{bmatrix} \frac{100}{3} & -\frac{25}{3} \\ -\frac{25}{3} & \frac{25}{3} \end{bmatrix}$

h. $\begin{bmatrix} 2 & 0 & 0 \\ 2 & 4 & 0 \\ 4 & 8 & 20 \end{bmatrix}$

1.7. Solve for matrix $\mathbf{X}$ in the matrix equation $\mathbf{AXB} + \mathbf{B} = \mathbf{AB}$. Assume that the inverses of $\mathbf{A}$ and $\mathbf{B}$ exist.

1.8. Solve each of the following for $\mathbf{x}$:

a.
$$\underset{\mathbf{C}}{\begin{bmatrix} 100/3 & -25/3 \\ -25/3 & 25/3 \end{bmatrix}} \underset{\mathbf{x}}{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}} \underset{=}{=} \underset{\mathbf{s}}{\begin{bmatrix} .01 \\ .11 \end{bmatrix}}$$

b.
$$\underset{\mathbf{C}}{\begin{bmatrix} .08 & .08 & .1 & 1 \\ .08 & .32 & .2 & 1 \\ .1 & .2 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}} \underset{\mathbf{x}}{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}} \underset{=}{=} \underset{\mathbf{s}}{\begin{bmatrix} .1 \\ .1 \\ .1 \\ .1 \end{bmatrix}}$$

1.9. Use matrices to solve the following system of equations:

$$.02x_1 + .04x_2 = .03$$
$$.06x_1 + .08x_2 = .01$$

1.10. Find the derivative of y with respect to x for the following polynomials:

a. $y = 7x^4$

b. $y = 5x^2 - 3x + 2$

c. $y = -7x^2 + 4x + 5$

1.11. **a.** At what value for x is y minimized in Problem 1.9.b? How do we know that y is not maximized at this point?

b. At what value for x is y maximized in Problem 1.9.c? How do we know that y is not minimized at this point?

1.12. Suppose the amount of lumber (stumpage value, the value of mature timber before it is cut) that could be produced from the timber in a given forest is a function of time, where s is the amount that can be produced and t is the number of years from today: $s(t) = t^3 - 3t^2 + t + 10$. This function reflects a recent fungus infection in many trees, and this fungus infection is expected to grow.

a. Find the (instantaneous) rate of change of its stumpage value 1 year from today. Verbally interpret your result.

b. Find the rate of change of its stumpage value 3 years after today.

c. Find the average rate of change of stumpage value from year 1 to year 3. Verbally interpret your result.

d. Suppose that the stumpage value function $s(t)$ reflects the fungus infection and the damage that it is likely to cause over the future. How might this damage be reflected in the value function?

1.13. Find derivatives for y with respect to x for each of the following:

a. $y = (4x + 2)^3$

b. $y = (3x^2 + 8)^{1/2}$

c. $y = 6x(4x^3 + 5x^2 + 3)$

d. $y = (1.5x - 4)^3(2.5x - 3.5)^4$

e. $y = 25/x^2$

f. $y = (6x - 16) \div (10x - 14)$

1.14. Use the following definition of a derivative (a) and the following statement (b) based on the binomial theorem to verify the power rule, also given below (c):

a. $$\frac{dy}{dx} = f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

b. $$(x + \Delta x)^n = \binom{n}{0} x^n(\Delta x)^0 + \binom{n}{1} x^{n-1}(\Delta x)^1 + \binom{n}{2} x^{n-2}(\Delta x)^2 + \ldots + \binom{n}{n-1} x^1(\Delta x)^{n-1} + \binom{n}{n} x^0(\Delta x)^n$$

c. If $y = \sum_{j=1}^{m} c_j \cdot x^{n_j}$, then $\frac{dy}{dx} = \sum_{j=1}^{m} c_j \cdot n_j \cdot x^{n_j - 1}$

1.15. Let $y = x^3$ and $x = t^2 + 1$. Use the chain rule to find dy/dt.

1.16. Differentiate each of the following with respect to x:

a. $y = e^{.05x}$
b. $y = (e^x)/x$
c. $y = 5\ln(x)$
d. $y = e^x \ln(x)$
e. $y = x^2 e^x$
f. $y = \ln(5x^3 + x)$
g. $y = 5x^3 - 6\sqrt{x} + 2e^x$
h. $y = x^2 \ln x$

1.17. Here is an exercise unrelated to finance. A square floor is measured to have side length 20 feet, with an error of plus or minus 0.1 feet. Use the differential to estimate the resulting possible error in measuring the area of the floor.

1.18. **a.** Consider the function $y = x^3$. Let $x_0 = 5$. Now, suppose we wish to increase x by $\Delta x = 1$ to $x_1 = 6$. Estimate y_1 based on a third-order Taylor approximation.
b. How does this approximation compare to an exact solution for y_1? Why?
c. Estimate y_1 based on a second-order Taylor approximation.
d. Estimate y_1 based on a first-order Taylor approximation.
e. Consider the function $y = 10x^3$. Let $x_0 = 2$, and suppose that we wish to increase x by $\Delta x = 3$ to $x_1 = 5$. Use first-, second-, then third-order Taylor polynomial expansions to evaluate y_1.

1.19. Our objective is to find the value for x, which enables us to maximize the function $y = 50x^2 - 10x$ subject to the constraint that $.1x \leq 100$. Set up and solve an appropriate Lagrange function for this problem. This exercise is intended to be a somewhat trivial illustration for setting up and solving a Lagrange optimization problem.

1.20. An investor wishes to budget her wealth $w = \$10{,}000$ in savings so that her spending over 4 years yields the highest level of utility (U, which can be considered to be satisfaction). She has mapped out a utility function that accounts for her consumption (x_t) each year t over the 4-year period:

$$U = 100x_1 + 200x_2 + 250x_3 + 350x_4 - .01x_1^2 - .2x_2^2 - .03x_3^2 - .04x_4^2 - .2x_3x_4$$

Unfortunately, price levels are expected to rise each year such that what \$1 buys in 1 year will cost \$2 in 2 years, \$3 in 3 years, and \$4 in 4 years. Her spending over the 4-year period cannot exceed \$10,000.

a. If this investor seeks to maximize her total utility over the 4-year period, what are optimal annual consumption levels for each year? Do bear in mind her \$10,000 wealth constraint.
b. What is the total utility level for the consumer?

1.21. **a.** Find the antiderivative for the function $f(x) = 10x - x^2$.
b. What is the area under the curve $f(x) = 10x - x^2$ between 0 and 1?
c. Find the Reimann sum for the function $f(x) = 10x - x^2$ based on five rectangles over the range 0 to 1.
d. Find the Reimann sum for the function $f(x) = 10x - x^2$ based on ten rectangles over the range 0 to 1.

1.22. Integrate each of the following functions over x:

a. $f(x) = 0$
b. $f(x) = 7$
c. $f(x) = 2x$
d. $f(x) = 21x^2$
e. $f(x) = 21x^2 + 5$
f. $f(x) = e^x$
g. $f(x) = .5e^{.5x}$
h. $f(x) = 5^x \ln(5)$
i. $f(x) = 1/x$
j. $f(x) = 5/x + 3e^x + 4x^2 - x$

1.23. **a.** Use the fundamental theorem of integral calculus to find the area between $x = 0$ and $x = 1$ under the function $f(x) = 8x - 9x^2$.
b. Plot out on an appropriate graph 20 rectangles representing the rectangles for the Reimann sum for this function between $x = 0$ and $x = 1$.

1.24. Consider the function $f(x) = 10x - x^2$. The area under a curve represented by this function over the range from $x = a = 0$ to $x = b = 1$ can be computed with a limit of Riemann sums or through the process of antidifferentiation. Verify that as the number of rectangles used to compute the Riemann sums approaches infinity, and the widths of these rectangles approach zero, the limit of the Riemann sums and antidifferentiation will produce the same area.

1.25. Evaluate $\int_1^3 x^2\,dx$.

1.26. Suppose that z, y, and x are all functions of t such that:

$$\frac{1}{z}\frac{dz}{dt} = x\frac{dx}{dt} + 5\frac{dy}{dt}$$

Find z in terms of x and y.

NOTES

1. *Normative models*, proposing what "ought to be," are distinguished from *positive models* that predict "what will be."
2. If it is possible to multiply two matrices, they are said to be *conformable* for multiplication. Any matrix can be multiplied by a scalar, where the product is simply each element times the value of the scalar.
3. **Y** is defined similarly to **X**. We can use the Gauss–Jordan elimination procedure to show that $\boldsymbol{\alpha}^T = [0, 0, 0]$ is the only solution to this equation, such that the set is linearly independent.
4. Note here that $\alpha_1 = -v_1 - \frac{1}{2}v_2 + \frac{1}{2}v_3$, $\alpha_2 = \frac{7}{5}v_1 + \frac{3}{10}v_2 - \frac{1}{2}v_3$ and $\alpha_3 = -3v_1 + v_3$.
5. The *law of one price* states that securities or portfolios producing the same payoff structures must sell for the same price. Arbitrage opportunities do not exist when the law of one price holds.
6. $f''(x) = \dfrac{d^2y}{dx^2} = \dfrac{d^2 \ln x}{dx^2} = \dfrac{dx^{-1}}{dx} = -x^{-2}$.
7. In many finance problems, we will want to use inequalities as constraints. However, it is convenient to convert them to equalities for Lagrange functions.
8. If the constraint is not binding, λ will equal zero. If the constraint is binding in this example, $x_1 + .2x_2$ will equal 10. Since either the contents within the parentheses or the Lagrange multiplier will equal zero, the numerical value of the function that we have added to our original function to be optimized will be zero. Although the numerical value of our original function is unchanged by the supplement representing the constraint, its derivatives will be affected by the constraint.