Appendix A

Basic Mathematical Tools

This appendix covers some basic mathematics that are used in econometric analysis. We summarize various properties of the summation operator, study properties of linear and certain nonlinear equations, and review proportions and percentages. We also present some special functions that often arise in applied econometrics, including quadratic functions and the natural logarithm. The first four sections require only basic algebra skills. Section A-5 contains a brief review of differential calculus; although a knowledge of calculus is not necessary to understand most of the text, it is used in some end-of-chapter appendices and in several of the more advanced chapters in Part 3.

A-1 The Summation Operator and Descriptive Statistics

The **summation operator** is a useful shorthand for manipulating expressions involving the sums of many numbers, and it plays a key role in statistics and econometric analysis. If $\{x_i: i = 1, \ldots, n\}$ denotes a sequence of n numbers, then we write the sum of these numbers as

$$\sum_{i=1}^{n} x_i \equiv x_1 + x_2 + \ldots + x_n. \quad \text{[A.1]}$$

With this definition, the summation operator is easily shown to have the following properties:

Property Sum.1: For any constant c,

$$\sum_{i=1}^{n} c = nc. \quad \text{[A.2]}$$

Property Sum.2: For any constant c,

$$\sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i. \quad \text{[A.3]}$$

Property Sum.3: If $\{(x_i, y_i): i = 1, 2, \ldots, n\}$ is a set of n pairs of numbers, and a and b are constants, then

$$\sum_{i=1}^{n}(ax_i + by_i) = a\sum_{i=1}^{n}x_i + b\sum_{i=1}^{n}y_i. \quad \text{[A.4]}$$

It is also important to be aware of some things that *cannot* be done with the summation operator. Let $\{(x_i, y_i): i = 1, 2, \ldots, n\}$ again be a set of n pairs of numbers with $y_i \neq 0$ for each i. Then,

$$\sum_{i=1}^{n}(x_i/y_i) \neq \left(\sum_{i=1}^{n}x_i\right)\bigg/\left(\sum_{i=1}^{n}y_i\right).$$

In other words, the sum of the ratios is not the ratio of the sums. In the $n = 2$ case, the application of familiar elementary algebra also reveals this lack of equality: $x_1/y_1 + x_2/y_2 \neq (x_1 + x_2)/(y_1 + y_2)$. Similarly, the sum of the squares is not the square of the sum: $\sum_{i=1}^{n}x_i^2 \neq \left(\sum_{i=1}^{n}x_i\right)^2$, except in special cases. That these two quantities are not generally equal is easiest to see when $n = 2$: $x_1^2 + x_2^2 \neq (x_1 + x_2)^2 = x_1^2 + 2x_1x_2 + x_2^2$.

Given n numbers $\{x_i: i = 1, \ldots, n\}$, we compute their **average** or *mean* by adding them up and dividing by n:

$$\bar{x} = (1/n)\sum_{i=1}^{n}x_i. \quad \text{[A.5]}$$

When the x_i are a sample of data on a particular variable (such as years of education), we often call this the *sample average* (or *sample mean*) to emphasize that it is computed from a particular set of data. The sample average is an example of a **descriptive statistic**; in this case, the statistic describes the central tendency of the set of points x_i.

There are some basic properties about averages that are important to understand. First, suppose we take each observation on x and subtract off the average: $d_i \equiv x_i - \bar{x}$ (the "d" here stands for *deviation* from the average). Then, the sum of these deviations is always zero:

$$\sum_{i=1}^{n}d_i = \sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n}x_i - \sum_{i=1}^{n}\bar{x} = \sum_{i=1}^{n}x_i - n\bar{x} = n\bar{x} - n\bar{x} = 0.$$

We summarize this as

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0. \quad \text{[A.6]}$$

A simple numerical example shows how this works. Suppose $n = 5$ and $x_1 = 6$, $x_2 = 1$, $x_3 = -2$, $x_4 = 0$, and $x_5 = 5$. Then, $\bar{x} = 2$, and the demeaned sample is $\{4, -1, -4, -2, 3\}$. Adding these gives zero, which is just what equation (A.6) says.

In our treatment of regression analysis in Chapter 2, we need to know some additional algebraic facts involving deviations from sample averages. An important one is that the sum of squared deviations is the sum of the squared x_i minus n times the square of $\bar{x}$:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}x_i^2 - n(\bar{x})^2. \quad \text{[A.7]}$$

This can be shown using basic properties of the summation operator:

$$\begin{aligned}\sum_{i=1}^{n}(x_i - \bar{x})^2 &= \sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \sum_{i=1}^{n}x_i^2 - 2\bar{x}\sum_{i=1}^{n}x_i + n(\bar{x})^2 \\ &= \sum_{i=1}^{n}x_i^2 - 2n(\bar{x})^2 + n(\bar{x})^2 = \sum_{i=1}^{n}x_i^2 - n(\bar{x})^2.\end{aligned}$$

Given a data set on two variables, $\{(x_i, y_i): i = 1, 2, \ldots, n\}$, it can also be shown that

$$\begin{aligned}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^{n}x_i(y_i - \bar{y}) \\ &= \sum_{i=1}^{n}(x_i - \bar{x})y_i = \sum_{i=1}^{n}x_iy_i - n(\bar{x}\cdot\bar{y});\end{aligned} \quad \text{[A.8]}$$

this is a generalization of equation (A.7). (There, $y_i = x_i$ for all i.)

The average is the measure of central tendency that we will focus on in most of this text. However, it is sometimes informative to use the **median** (or *sample median*) to describe the central value. To obtain the median of the n numbers $\{x_1, \ldots, x_n\}$, we first order the values of the x_i from smallest to largest. Then, if n is odd, the sample median is the middle number of the ordered observations. For example, given the numbers $\{-4, 8, 2, 0, 21, -10, 18\}$, the median value is 2 (because the ordered sequence is $\{-10, -4, 0, 2, 8, 18, 21\}$). If we change the largest number in this list, 21, to twice its value, 42, the median is still 2. By contrast, the sample average would increase from 5 to 8, a sizable change. Generally, the median is less sensitive than the average to changes in the extreme values (large or small) in a list of numbers. This is why "median incomes" or "median housing values" are often reported, rather than averages, when summarizing income or housing values in a city or county.

If n is even, there is no unique way to define the median because there are two numbers at the center. Usually, the median is defined to be the average of the two middle values (again, after ordering the numbers from smallest to largest). Using this rule, the median for the set of numbers $\{4, 12, 2, 6\}$ would be $(4 + 6)/2 = 5$.

A-2 Properties of Linear Functions

Linear functions play an important role in econometrics because they are simple to interpret and manipulate. If x and y are two variables related by

$$y = \beta_0 + \beta_1 x, \quad \text{[A.9]}$$

then we say that y is a **linear function** of x, and β_0 and β_1 are two parameters (numbers) describing this relationship. The **intercept** is β_0, and the **slope** is β_1.

The defining feature of a linear function is that the change in y is always β_1 times the change in x:

$$\Delta y = \beta_1 \Delta x, \quad \text{[A.10]}$$

where Δ denotes "change." In other words, the **marginal effect** of x on y is constant and equal to β_1.

EXAMPLE A.1 Linear Housing Expenditure Function

Suppose that the relationship between monthly housing expenditure and monthly income is

$$housing = 164 + .27\ income. \qquad \text{[A.11]}$$

Then, for each additional dollar of income, 27 cents is spent on housing. If family income increases by \$200, then housing expenditure increases by $(.27)200 = \$54$. This function is graphed in Figure A.1.

According to equation (A.11), a family with no income spends \$164 on housing, which of course cannot be literally true. For low levels of income, this linear function would not describe the relationship between *housing* and *income* very well, which is why we will eventually have to use other types of functions to describe such relationships.

In (A.11), the *marginal propensity to consume* (MPC) housing out of income is .27. This is different from the *average propensity to consume* (APC), which is

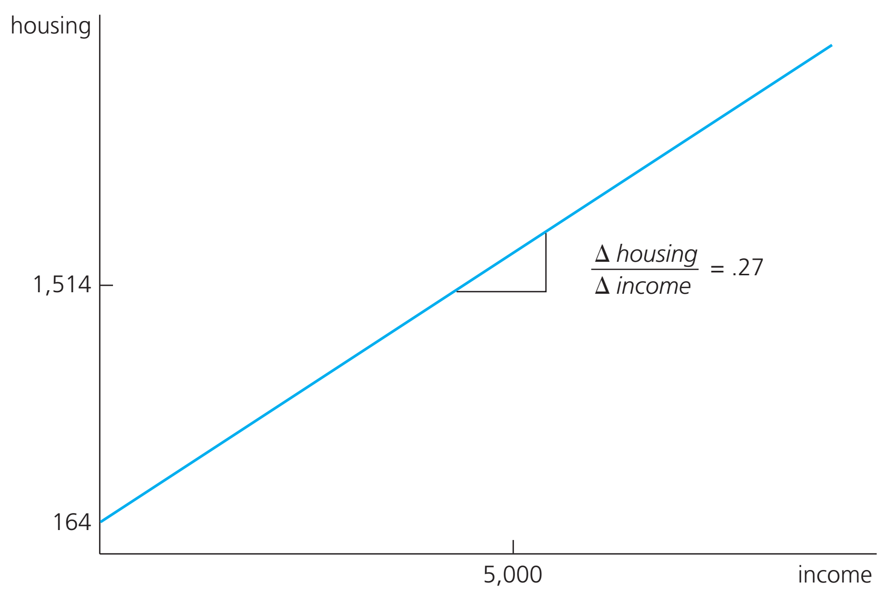
$$\frac{housing}{income} = 164/income + .27.$$

The APC is not constant; it is always larger than the MPC, and it gets closer to the MPC as income increases.

Linear functions are easily defined for more than two variables. Suppose that y is related to two variables, x_1 and x_2, in the general form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2. \qquad \text{[A.12]}$$

FIGURE A.1 Graph of *housing* = 164 + .27 *income*.



It is rather difficult to envision this function because its graph is three-dimensional. Nevertheless, β_0 is still the intercept (the value of y when $x_1 = 0$ and $x_2 = 0$), and β_1 and β_2 measure particular slopes. From (A.12), the change in y, for given changes in x_1 and x_2, is

$$\Delta y = \beta_1 \Delta x_1 + \beta_2 \Delta x_2. \quad \text{[A.13]}$$

If x_2 does not change, that is, $\Delta x_2 = 0$, then we have

$$\Delta y = \beta_1 \Delta x_1 \text{ if } \Delta x_2 = 0,$$

so that β_1 is the slope of the relationship in the direction of x_1:

$$\beta_1 = \frac{\Delta y}{\Delta x_1} \text{ if } \Delta x_2 = 0.$$

Because it measures how y changes with x_1, holding x_2 fixed, β_1 is often called the **partial effect** of x_1 on y. Because the partial effect involves holding other factors fixed, it is closely linked to the notion of **ceteris paribus**. The parameter β_2 has a similar interpretation: $\beta_2 = \Delta y / \Delta x_2$ if $\Delta x_1 = 0$, so that β_2 is the partial effect of x_2 on y.

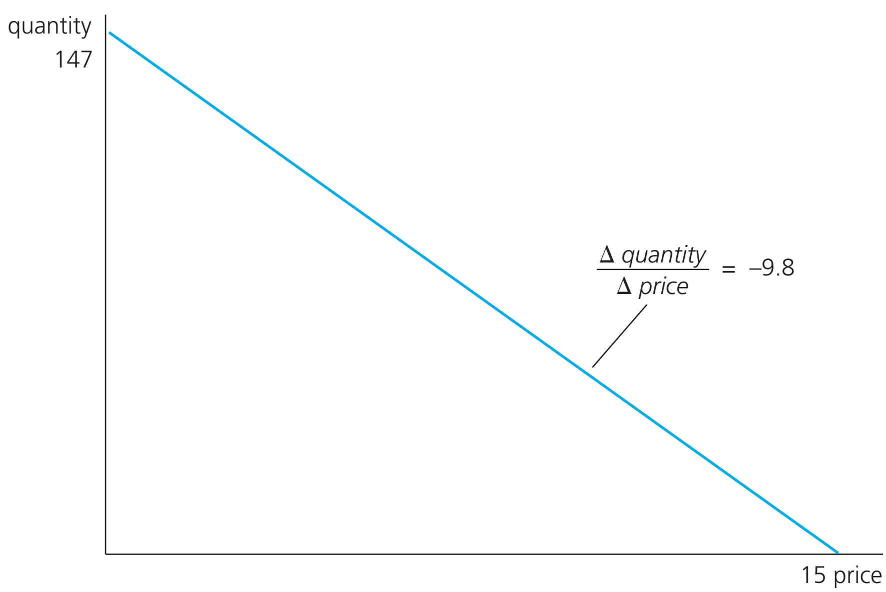
EXAMPLE A.2 Demand for Compact Discs

For college students, suppose that the monthly quantity demanded of compact discs is related to the price of compact discs and monthly discretionary income by

$$quantity = 120 - 9.8\, price + .03\, income,$$

where *price* is dollars per disc and *income* is measured in dollars. The *demand curve* is the relationship between *quantity* and *price*, holding *income* (and other factors) fixed. This is graphed in two dimensions in Figure A.2 at an income level of \$900. The slope of the demand curve, -9.8, is the *partial effect* of price on quantity: holding income fixed, if the price of compact discs increases by one dollar, then the quantity demanded falls by 9.8. (We abstract from the fact that CDs can only be purchased in discrete units.) An increase in income simply shifts the demand curve up (changes the intercept), but the slope remains the same.

FIGURE A.2 Graph of $quantity = 120 - 9.8\, price + .03\, income$, with income fixed at \$900.



A-3 Proportions and Percentages

Proportions and percentages play such an important role in applied economics that it is necessary to become very comfortable in working with them. Many quantities reported in the popular press are in the form of percentages; a few examples are interest rates, unemployment rates, and high school graduation rates.

An important skill is being able to convert proportions to percentages and vice versa. A percentage is easily obtained by multiplying a proportion by 100. For example, if the proportion of adults in a county with a high school degree is .82, then we say that 82% (82 percent) of adults have a high school degree. Another way to think of percentages and proportions is that a proportion is the decimal form of a percentage. For example, if the marginal tax rate for a family earning $30,000 per year is reported as 28%, then the proportion of the next dollar of income that is paid in income taxes is .28 (or 28¢).

When using percentages, we often need to convert them to decimal form. For example, if a state sales tax is 6% and $200 is spent on a taxable item, then the sales tax paid is 200(.06) = $12. If the annual return on a certificate of deposit (CD) is 7.6% and we invest $3,000 in such a CD at the beginning of the year, then our interest income is 3,000(.076) = $228. As much as we would like it, the interest income is not obtained by multiplying 3,000 by 7.6.

We must be wary of proportions that are sometimes incorrectly reported as percentages in the popular media. If we read, "The percentage of high school students who drink alcohol is .57," we know that this really means 57% (not just over one-half of a percent, as the statement literally implies). College volleyball fans are probably familiar with press clips containing statements such as "Her hitting percentage was .372." This really means that her hitting percentage was 37.2%.

In econometrics, we are often interested in measuring the *changes* in various quantities. Let x denote some variable, such as an individual's income, the number of crimes committed in a community, or the profits of a firm. Let x_0 and x_1 denote two values for x: x_0 is the initial value, and x_1 is the subsequent value. For example, x_0 could be the annual income of an individual in 1994 and x_1 the income of the same individual in 1995. The **proportionate change** in x in moving from x_0 to x_1, sometimes called the **relative change,** is simply

$$(x_1 - x_0)/x_0 = \Delta x/x_0, \quad \text{[A.14]}$$

assuming, of course, that $x_0 \neq 0$. In other words, to get the proportionate change, we simply divide the change in x by its initial value. This is a way of standardizing the change so that it is free of units. For example, if an individual's income goes from $30,000 per year to $36,000 per year, then the proportionate change is 6,000/30,000 = .20.

It is more common to state changes in terms of percentages. The **percentage change** in x in going from x_0 to x_1 is simply 100 times the proportionate change:

$$\%\Delta x = 100(\Delta x/x_0); \quad \text{[A.15]}$$

the notation "$\%\Delta x$" is read as "the percentage change in x." For example, when income goes from $30,000 to $33,750, income has increased by 12.5%; to get this, we simply multiply the proportionate change, .125, by 100.

Again, we must be on guard for proportionate changes that are reported as percentage changes. In the previous example, for instance, reporting the percentage change in income as .125 is incorrect and could lead to confusion.

When we look at changes in things like dollar amounts or population, there is no ambiguity about what is meant by a percentage change. By contrast, interpreting percentage change calculations can be tricky when the variable of interest is itself a percentage, something that happens often in economics and other social sciences. To illustrate, let x denote the percentage of adults in a particular city having a college education. Suppose the initial value is $x_0 = 24$ (24% have a college education), and the new value is $x_1 = 30$. We can compute two quantities to describe how the percentage of college-educated people has changed. The first is the change in x, Δx. In this case, $\Delta x = x_1 - x_0 = 6$: the percentage of people with a college education has increased by six *percentage points*. On the other hand, we can compute the percentage change in x using equation (A.15): $\%\Delta x = 100[(30 - 24)/24] = 25$.

In this example, the percentage point change and the percentage change are very different. The **percentage point change** is just the change in the percentages. The percentage change is the change relative to the initial value. Generally, we must pay close attention to which number is being computed. The careful researcher makes this distinction perfectly clear; unfortunately, in the popular press as well as in academic research, the type of reported change is often unclear.

EXAMPLE A.3 **Michigan Sales Tax Increase**

In March 1994, Michigan voters approved a sales tax increase from 4% to 6%. In political advertisements, supporters of the measure referred to this as a two percentage point increase, or an increase of two cents on the dollar. Opponents of the tax increase called it a 50% increase in the sales tax rate. Both claims are correct; they are simply different ways of measuring the increase in the sales tax. Naturally, each group reported the measure that made its position most favorable.

For a variable such as salary, it makes no sense to talk of a "percentage point change in salary" because salary is not measured as a percentage. We can describe a change in salary either in dollar or percentage terms.

A-4 Some Special Functions and Their Properties

In Section A-2, we reviewed the basic properties of linear functions. We already indicated one important feature of functions like $y = \beta_0 + \beta_1 x$: a one-unit change in x results in the *same* change in y, regardless of the initial value of x. As we noted earlier, this is the same as saying the marginal effect of x on y is constant, something that is not realistic for many economic relationships. For example, the important economic notion of *diminishing marginal returns* is not consistent with a linear relationship.

In order to model a variety of economic phenomena, we need to study several nonlinear functions. A **nonlinear function** is characterized by the fact that the change in y for a given change in x depends on the starting value of x. Certain nonlinear functions appear frequently in empirical economics, so it is important to know how to interpret them. A complete understanding of nonlinear functions takes us into the realm of calculus. Here, we simply summarize the most significant aspects of the functions, leaving the details of some derivations for Section A-5.

A-4a Quadratic Functions

One simple way to capture diminishing returns is to add a quadratic term to a linear relationship. Consider the equation

$$y = \beta_0 + \beta_1 x + \beta_2 x^2, \quad \text{[A.16]}$$

where β_0, β_1, and β_2 are parameters. When $\beta_1 > 0$ and $\beta_2 < 0$, the relationship between y and x has the parabolic shape given in Figure A.3, where $\beta_0 = 6$, $\beta_1 = 8$, and $\beta_2 = -2$.

When $\beta_1 > 0$ and $\beta_2 < 0$, it can be shown (using calculus in the next section) that the *maximum* of the function occurs at the point

$$x^* = \beta_1/(-2\beta_2). \quad \text{[A.17]}$$

For example, if $y = 6 + 8x - 2x^2$ (so $\beta_1 = 8$ and $\beta_2 = -2$), then the largest value of y occurs at $x^* = 8/4 = 2$, and this value is $6 + 8(2) - 2(2)^2 = 14$ (see Figure A.3).

The fact that equation (A.16) implies a **diminishing marginal effect** of x on y is easily seen from its graph. Suppose we start at a low value of x and then increase x by some amount, say, c. This has a larger effect on y than if we start at a higher value of x and increase x by the same amount c. In fact, once $x > x^*$, an increase in x actually decreases y.

The statement that x has a diminishing marginal effect on y is the same as saying that the slope of the function in Figure A.3 decreases as x increases. Although this is clear from looking at the graph, we usually want to quantify how quickly the slope is changing. An application of calculus gives the approximate slope of the quadratic function as

$$slope = \frac{\Delta y}{\Delta x} \approx \beta_1 + 2\beta_2 x, \quad \text{[A.18]}$$

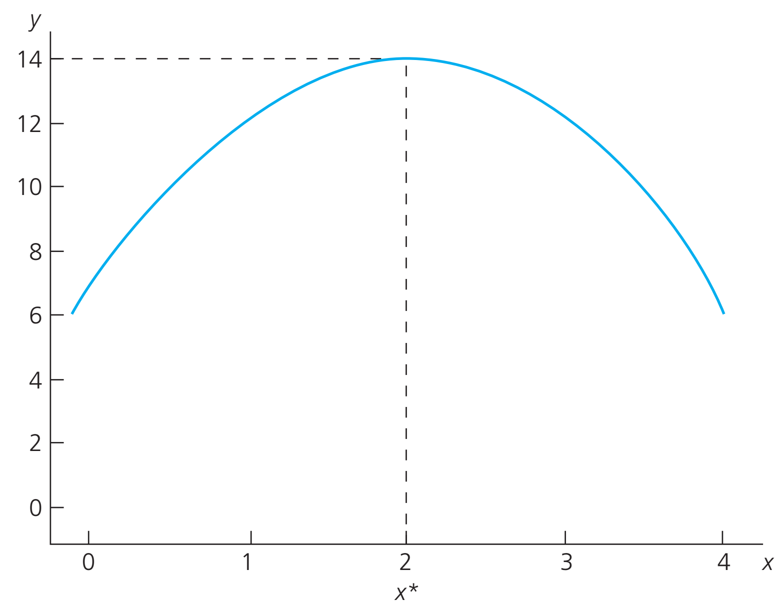
for "small" changes in x. [The right-hand side of equation (A.18) is the **derivative** of the function in equation (A.16) with respect to x.] Another way to write this is

$$\Delta y \approx (\beta_1 + 2\beta_2 x)\Delta x \text{ for "small" } \Delta x. \quad \text{[A.19]}$$

To see how well this approximation works, consider again the function $y = 6 + 8x - 2x^2$. Then, according to equation (A.19), $\Delta y \approx (8 - 4x)\Delta x$. Now, suppose we start at $x = 1$ and change x by $\Delta x = .1$. Using (A.19), $\Delta y \approx (8 - 4)(.1) = .4$. Of course, we can compute the change exactly by finding the values of y when $x = 1$ and $x = 1.1$: $y_0 = 6 + 8(1) - 2(1)^2 = 12$ and $y_1 = 6 + 8(1.1) - 2(1.1)^2 = 12.38$, so the exact change in y is .38. The approximation is pretty close in this case.

Now, suppose we start at $x = 1$ but change x by a larger amount: $\Delta x = .5$. Then, the approximation gives $\Delta y \approx 4(.5) = 2$. The exact change is determined by finding the difference in y when $x = 1$

FIGURE A.3 Graph of $y = 6 + 8x - 2x^2$.



and $x = 1.5$. The former value of y was 12, and the latter value is $6 + 8(1.5) - 2(1.5)^2 = 13.5$, so the actual change is 1.5 (not 2). The approximation is worse in this case because the change in x is larger.

For many applications, equation (A.19) can be used to compute the approximate marginal effect of x on y for any initial value of x and small changes. And, we can always compute the exact change if necessary.

EXAMPLE A.4 A Quadratic Wage Function

Suppose the relationship between hourly wages and years in the workforce (*exper*) is given by

$$wage = 5.25 + .48\ exper - .008\ exper^2. \quad \text{[A.20]}$$

This function has the same general shape as the one in Figure A.3. Using equation (A.17), *exper* has a positive effect on wage up to the turning point, $exper^* = .48/[2(.008)] = 30$. The first year of experience is worth approximately .48, or 48 cents [see (A.19) with $x = 0$, $\Delta x = 1$]. Each additional year of experience increases wage by less than the previous year—reflecting a diminishing marginal return to experience. At 30 years, an additional year of experience would actually lower the wage. This is not very realistic, but it is one of the consequences of using a quadratic function to capture a diminishing marginal effect: at some point, the function must reach a maximum and curve downward. For practical purposes, the point at which this happens is often large enough to be inconsequential, but not always.

The graph of the quadratic function in (A.16) has a U-shape if $\beta_1 < 0$ and $\beta_2 > 0$, in which case there is an increasing marginal return. The minimum of the function is at the point $-\beta_1/(2\beta_2)$.

A-4b The Natural Logarithm

The nonlinear function that plays the most important role in econometric analysis is the **natural logarithm**. In this text, we denote the natural logarithm, which we often refer to simply as the **log function**, as

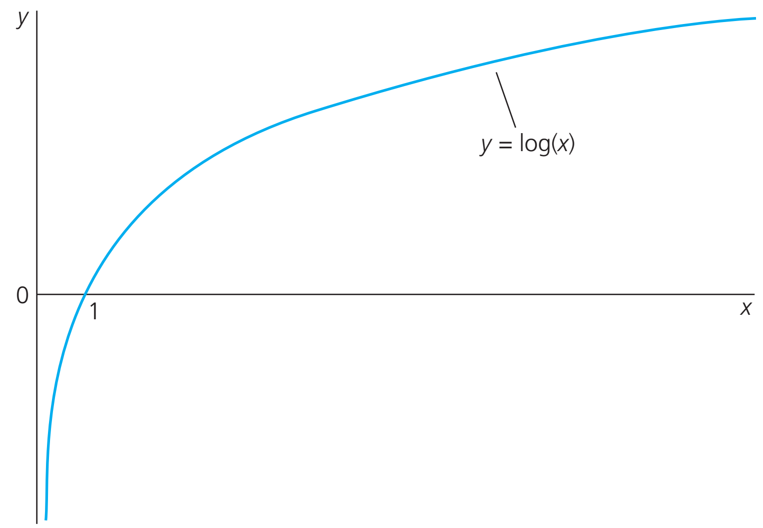
$$y = \log(x). \quad \text{[A.21]}$$

You might remember learning different symbols for the natural log; $\ln(x)$ or $\log_e(x)$ are the most common. These different notations are useful when logarithms with several different bases are being used. For our purposes, only the natural logarithm is important, and so $\log(x)$ denotes the natural logarithm throughout this text. This corresponds to the notational usage in many statistical packages, although some use $\ln(x)$ [and most calculators use $\ln(x)$]. Economists use both $\log(x)$ and $\ln(x)$, which is useful to know when you are reading papers in applied economics.

The function $y = \log(x)$ is defined only for $x > 0$, and it is plotted in Figure A.4. It is not very important to know how the values of $\log(x)$ are obtained. For our purposes, the function can be thought of as a black box: we can plug in any $x > 0$ and obtain $\log(x)$ from a calculator or a computer.

Several things are apparent from Figure A.4. First, when $y = \log(x)$, the relationship between y and x displays diminishing marginal returns. One important difference between the log and the quadratic function in Figure A.3 is that when $y = \log(x)$, the effect of x on y never becomes negative: the slope of the function gets closer and closer to zero as x gets large, but the slope never quite reaches zero and certainly never becomes negative.

FIGURE A.4 Graph of $y = \log(x)$.



The following are also apparent from Figure A.4:

$$\log(x) < 0 \text{ for } 0 < x < 1$$
$$\log(1) = 0$$
$$\log(x) > 0 \text{ for } x > 1.$$

In particular, $\log(x)$ can be positive or negative. Some useful algebraic facts about the log function are

$$\log(x_1 \cdot x_2) = \log(x_1) + \log(x_2),\ x_1, x_2 > 0$$
$$\log(x_1/x_2) = \log(x_1) - \log(x_2),\ x_1, x_2 > 0$$
$$\log(x^c) \quad = c\log(x),\ x > 0,\ c \text{ any number.}$$

Occasionally, we will need to rely on these properties.

The logarithm can be used for various approximations that arise in econometric applications. First, $\log(1 + x) \approx x$ for $x \approx 0$. You can try this with $x = .02$, .1, and .5 to see how the quality of the approximation deteriorates as x gets larger. Even more useful is the fact that the difference in logs can be used to approximate proportionate changes. Let x_0 and x_1 be positive values. Then, it can be shown (using calculus) that

$$\log(x_1) - \log(x_0) \approx (x_1 - x_0)/x_0 = \Delta x/x_0 \qquad \text{[A.22]}$$

for small changes in x. If we multiply equation (A.22) by 100 and write $\Delta\log(x) = \log(x_1) - \log(x_0)$, then

$$100 \cdot \Delta\log(x) \approx \%\Delta x \qquad \text{[A.23]}$$

for small changes in x. The meaning of "small" depends on the context, and we will encounter several examples throughout this text.

Why should we approximate the percentage change using (A.23) when the exact percentage change is so easy to compute? Momentarily, we will see why the approximation in (A.23) is useful in econometrics. First, let us see how good the approximation is in two examples.

First, suppose $x_0 = 40$ and $x_1 = 41$. Then, the percentage change in x in moving from x_0 to x_1 is 2.5%, using $100(x_1 - x_0)/x_0$. Now, $\log(41) - \log(40) = .0247$ (to four decimal places), which when multiplied by 100 is very close to 2.5. The approximation works pretty well. Now, consider a much bigger change: $x_0 = 40$ and $x_1 = 60$. The exact percentage change is 50%. However, $\log(60) - \log(40) \approx .4055$, so the approximation gives 40.55%, which is much farther off.

Why is the approximation in (A.23) useful if it is only satisfactory for small changes? To build up to the answer, we first define the **elasticity** of y with respect to x as

$$\frac{\Delta y}{\Delta x} \cdot \frac{x}{y} = \frac{\%\Delta y}{\%\Delta x}. \qquad \text{[A.24]}$$

In other words, the elasticity of y with respect to x is the percentage change in y when x increases by 1%. This notion should be familiar from introductory economics.

If y is a linear function of x, $y = \beta_0 + \beta_1 x$, then the elasticity is

$$\frac{\Delta y}{\Delta x} \cdot \frac{x}{y} = \beta_1 \cdot \frac{x}{y} = \beta_1 \cdot \frac{x}{\beta_0 + \beta_1 x}, \qquad \text{[A.25]}$$

which clearly depends on the value of x. (This is a generalization of the well-known result from basic demand theory: the elasticity is not constant along a straight-line demand curve.)

Elasticities are of critical importance in many areas of applied economics, not just in demand theory. It is convenient in many situations to have *constant* elasticity models, and the log function allows us to specify such models. If we use the approximation in (A.23) for both x and y, then the elasticity is approximately equal to $\Delta\log(y)/\Delta\log(x)$. Thus, a constant elasticity model is approximated by the equation

$$\log(y) = \beta_0 + \beta_1 \log(x), \qquad \text{[A.26]}$$

and β_1 is the elasticity of y with respect to x (assuming that $x, y > 0$).

EXAMPLE A.5 Constant Elasticity Demand Function

If q is quantity demanded and p is price and these variables are related by

$$\log(q) = 4.7 - 1.25 \log(p),$$

then the price elasticity of demand is -1.25. Roughly, a 1% increase in price leads to a 1.25% fall in the quantity demanded.

For our purposes, the fact that β_1 in (A.26) is only close to the elasticity is not important. In fact, when the elasticity is defined using calculus—as in Section A-5—the definition is exact. For the purposes of econometric analysis, (A.26) defines a **constant elasticity model**. Such models play a large role in empirical economics.

Other possibilities for using the log function often arise in empirical work. Suppose that $y > 0$ and

$$\log(y) = \beta_0 + \beta_1 x. \qquad \text{[A.27]}$$

Then, $\Delta\log(y) = \beta_1 \Delta x$, so $100 \cdot \Delta\log(y) = (100 \cdot \beta_1)\Delta x$. It follows that, when y and x are related by equation (A.27),

$$\%\Delta y \approx (100 \cdot \beta_1)\Delta x. \qquad \text{[A.28]}$$

EXAMPLE A.6 **Logarithmic Wage Equation**

Suppose that hourly wage and years of education are related by

$$\log(wage) = 2.78 + .094\ educ.$$

Then, using equation (A.28),

$$\%\Delta wage \approx 100(.094)\Delta educ = 9.4\ \Delta educ.$$

It follows that one more year of education increases hourly wage by about 9.4%.

Generally, the quantity $\%\Delta y/\Delta x$ is called the **semi-elasticity** of y with respect to x. The semi-elasticity is the percentage change in y when x increases by one *unit*. What we have just shown is that, in model (A.27), the semi-elasticity is constant and equal to $100\cdot\beta_1$. In Example A.6, we can conveniently summarize the relationship between wages and education by saying that one more year of education—starting from any amount of education—increases the wage by about 9.4%. This is why such models play an important role in economics.

Another relationship of some interest in applied economics is

$$y = \beta_0 + \beta_1\log(x), \qquad [A.29]$$

where $x > 0$. How can we interpret this equation? If we take the change in y, we get $\Delta y = \beta_1\Delta\log(x)$, which can be rewritten as $\Delta y = (\beta_1/100)[100\cdot\Delta\log(x)]$. Thus, using the approximation in (A.23), we have

$$\Delta y \approx (\beta_1/100)(\%\Delta x). \qquad [A.30]$$

In other words, $\beta_1/100$ is the unit change in y when x increases by 1%.

EXAMPLE A.7 **Labor Supply Function**

Assume that the labor supply of a worker can be described by

$$hours = 33 + 45.1\ \log(wage),$$

where *wage* is hourly wage and *hours* is hours worked per week. Then, from (A.30),

$$\Delta hours \approx (45.1/100)(\%\Delta wage) = .451\ \%\Delta wage.$$

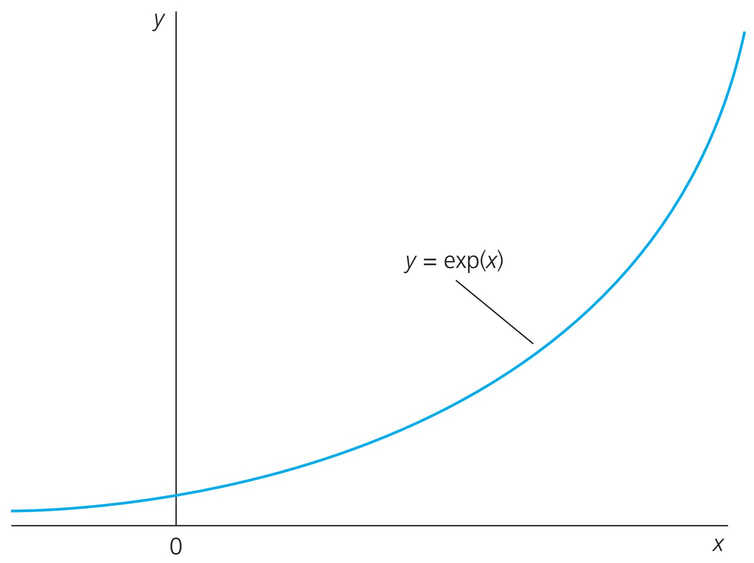
In other words, a 1% increase in *wage* increases the weekly hours worked by about .45, or slightly less than one-half hour. If the wage increases by 10%, then $\Delta hours = .451(10) = 4.51$, or about four and one-half hours. We would not want to use this approximation for much larger percentage changes in wages.

A-4c The Exponential Function

Before leaving this section, we need to discuss a special function that is related to the log. As motivation, consider equation (A.27). There, $\log(y)$ is a linear function of x. But how do we find y itself as a function of x? The answer is given by the **exponential function**.

We will write the exponential function as $y = \exp(x)$, which is graphed in Figure A.5. From Figure A.5, we see that $\exp(x)$ is defined for any value of x and is always greater than zero. Sometimes, the exponential function is written as $y = e^x$, but we will not use this notation. Two important values of the exponential function are $\exp(0) = 1$ and $\exp(1) = 2.7183$ (to four decimal places).

FIGURE A.5 Graph of $y = \exp(x)$.



The exponential function is the inverse of the log function in the following sense: $\log[\exp(x)] = x$ for all x, and $\exp[\log(x)] = x$ for $x > 0$. In other words, the log "undoes" the exponential, and vice versa. (This is why the exponential function is sometimes called the *anti-log* function.) In particular, note that $\log(y) = \beta_0 + \beta_1 x$ is equivalent to

$$y = \exp(\beta_0 + \beta_1 x).$$

If $\beta_1 > 0$, the relationship between x and y has the same shape as in Figure A.5. Thus, if $\log(y) = \beta_0 + \beta_1 x$ with $\beta_1 > 0$, then x has an *increasing* marginal effect on y. In Example A.6, this means that another year of education leads to a larger change in wage than the previous year of education.

Two useful facts about the exponential function are $\exp(x_1 + x_2) = \exp(x_1)\exp(x_2)$ and $\exp[c \cdot \log(x)] = x^c$.

A-5 Differential Calculus

In the previous section, we asserted several approximations that have foundations in calculus. Let $y = f(x)$ for some function f. Then, for small changes in x,

$$\Delta y \approx \frac{df}{dx} \cdot \Delta x, \qquad \text{[A.31]}$$

where df/dx is the derivative of the function f, evaluated at the initial point x_0. We also write the derivative as dy/dx.

For example, if $y = \log(x)$, then $dy/dx = 1/x$. Using (A.31), with dy/dx evaluated at x_0, we have $\Delta y \approx (1/x_0)\Delta x$, or $\Delta \log(x) \approx \Delta x/x_0$, which is the approximation given in (A.22).

In applying econometrics, it helps to recall the derivatives of a handful of functions because we use the derivative to define the slope of a function at a given point. We can then use (A.31) to find the approximate change in y for small changes in x. In the linear case, the derivative is simply the slope of the line, as we would hope: if $y = \beta_0 + \beta_1 x$, then $dy/dx = \beta_1$.

If $y = x^c$, then $dy/dx = cx^{c-1}$. The derivative of a sum of two functions is the sum of the derivatives: $d[f(x) + g(x)]/dx = df(x)/dx + dg(x)/dx$. The derivative of a constant times any function is that same constant times the derivative of the function: $d[cf(x)]/dx = c[df(x)/dx]$. These simple rules allow us to find derivatives of more complicated functions. Other rules, such as the product, quotient, and chain rules, will be familiar to those who have taken calculus, but we will not review those here.

Some functions that are often used in economics, along with their derivatives, are

$$y = \beta_0 + \beta_1 x + \beta_2 x^2;\ dy/dx = \beta_1 + 2\beta_2 x$$
$$y = \beta_0 + \beta_1/x;\ dy/dx = -\beta_1/(x^2)$$
$$y = \beta_0 + \beta_1\sqrt{x};\ dy/dx = (\beta_1/2)x^{-1/2}$$
$$y = \beta_0 + \beta_1 \log(x);\ dy/dx = \beta_1/x$$
$$y = \exp(\beta_0 + \beta_1 x);\ dy/dx = \beta_1 \exp(\beta_0 + \beta_1 x).$$

If $\beta_0 = 0$ and $\beta_1 = 1$ in this last expression, we get $dy/dx = \exp(x)$, when $y = \exp(x)$.

In Section A-4, we noted that equation (A.26) defines a constant elasticity model when calculus is used. The calculus definition of elasticity is $(dy/dx)\cdot(x/y)$. It can be shown using properties of logs and exponentials that, when (A.26) holds, $(dy/dx)\cdot(x/y) = \beta_1$.

When y is a function of multiple variables, the notion of a **partial derivative** becomes important. Suppose that

$$y = f(x_1, x_2). \quad \text{[A.32]}$$

Then, there are two partial derivatives, one with respect to x_1 and one with respect to x_2. The partial derivative of y with respect to x_1, denoted here by $\partial y/\partial x_1$, is just the usual derivative of (A.32) with respect to x_1, where x_2 is treated as a *constant*. Similarly, $\partial y/\partial x_2$ is just the derivative of (A.32) with respect to x_2, holding x_1 fixed.

Partial derivatives are useful for much the same reason as ordinary derivatives. We can approximate the change in y as

$$\Delta y \approx \frac{\partial y}{\partial x_1}\cdot \Delta x_1, \text{ holding } x_2 \text{ fixed.} \quad \text{[A.33]}$$

Thus, calculus allows us to define partial effects in nonlinear models just as we could in linear models. In fact, if

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2,$$

then

$$\frac{\partial y}{\partial x_1} = \beta_1, \frac{\partial y}{\partial x_2} = \beta_2.$$

These can be recognized as the partial effects defined in Section A-2.

A more complicated example is

$$y = 5 + 4x_1 + x_1^2 - 3x_2 + 7x_1 \cdot x_2. \quad \text{[A.34]}$$

Now, the derivative of (A.34), with respect to x_1 (treating x_2 as a constant), is simply

$$\frac{\partial y}{\partial x_1} = 4 + 2x_1 + 7x_2;$$

note how this depends on x_1 and x_2. The derivative of (A.34), with respect to x_2, is $\partial y/\partial x_2 = -3 + 7x_1$, so this depends only on x_1.

EXAMPLE A.8 Wage Function with Interaction

A function relating wages to years of education and experience is

$$wage = 3.10 + .41\ educ + .19\ exper - .004\ exper^2 + .007\ educ \cdot exper. \quad \text{[A.35]}$$

The partial effect of *exper* on *wage* is the partial derivative of (A.35):

$$\frac{\partial wage}{\partial exper} = .19 - .008\ exper + .007\ educ.$$

This is the approximate change in wage due to increasing experience by one year. Notice that this partial effect depends on the initial level of *exper* and *educ*. For example, for a worker who is starting with $educ = 12$ and $exper = 5$, the next year of experience increases wage by about $.19 - .008(5) + .007(12) = .234$, or 23.4 cents per hour. The exact change can be calculated by computing (A.35) at $exper = 5$, $educ = 12$ and at $exper = 6$, $educ = 12$, and then taking the difference. This turns out to be .23, which is very close to the approximation.

Differential calculus plays an important role in minimizing and maximizing functions of one or more variables. If $f(x_1, x_2, \ldots, x_k)$ is a differentiable function of k variables, then a necessary condition for $x_1^*, x_2^*, \ldots, x_k^*$ to either minimize or maximize f over all possible values of x_j is

$$\frac{\partial f}{\partial x_i}(x_1^*, x_2^*, \ldots, x_k^*) = 0, j = 1, 2, \ldots, k. \quad \text{[A.36]}$$

In other words, all of the partial derivatives of f must be zero when they are evaluated at the x_h*. These are called the *first order conditions* for minimizing or maximizing a function. Practically, we hope to solve equation (A.36) for the x_h*. Then, we can use other criteria to determine whether we have minimized or maximized the function. We will not need those here. [See Sydsaeter and Hammond (1995) for a discussion of multivariable calculus and its use in optimizing functions.]

Summary

The math tools reviewed here are crucial for understanding regression analysis and the probability and statistics that are covered in Appendices B and C. The material on nonlinear functions—especially quadratic, logarithmic, and exponential functions—is critical for understanding modern applied economic research. The level of comprehension required of these functions does not include a deep knowledge of calculus, although calculus is needed for certain derivations.

Key Terms

Average
Ceteris Paribus
Constant Elasticity Model
Derivative
Descriptive Statistic
Diminishing Marginal Effect
Elasticity
Exponential Function
Intercept
Linear Function
Log Function
Marginal Effect
Median
Natural Logarithm
Nonlinear Function
Partial Derivative
Partial Effect
Percentage Change
Percentage Point Change
Proportionate Change
Relative Change
Semi-Elasticity
Slope
Summation Operator

Problems

1 The following table contains monthly housing expenditures for 10 families.

Family	Monthly Housing Expenditures (Dollars)
1	300
2	440
3	350
4	1,100
5	640
6	480
7	450
8	700
9	670
10	530

(i) Find the average monthly housing expenditure.
(ii) Find the median monthly housing expenditure.
(iii) If monthly housing expenditures were measured in hundreds of dollars, rather than in dollars, what would be the average and median expenditures?
(iv) Suppose that family number 8 increases its monthly housing expenditure to $900, but the expenditures of all other families remain the same. Compute the average and median housing expenditures.

2 Suppose the following equation describes the relationship between the average number of classes missed during a semester (*missed*) and the distance from school (*distance*, measured in miles):

$$missed = 3 + 0.2\ distance.$$

(i) Sketch this line, being sure to label the axes. How do you interpret the intercept in this equation?
(ii) What is the average number of classes missed for someone who lives five miles away?
(iii) What is the difference in the average number of classes missed for someone who lives 10 miles away and someone who lives 20 miles away?

3 In Example A.2, quantity of compact discs was related to price and income by $quantity = 120 - 9.8\ price + .03\ income$. What is the demand for CDs if $price = 15$ and $income = 200$? What does this suggest about using linear functions to describe demand curves?

4 Suppose the unemployment rate in the United States goes from 6.4% in one year to 5.6% in the next.
(i) What is the percentage point decrease in the unemployment rate?
(ii) By what percentage has the unemployment rate fallen?

5 Suppose that the return from holding a particular firm's stock goes from 15% in one year to 18% in the following year. The majority shareholder claims that "the stock return only increased by 3%," while the chief executive officer claims that "the return on the firm's stock increased by 20%." Reconcile their disagreement.

6 Suppose that Person A earns $35,000 per year and Person B earns $42,000.
(i) Find the exact percentage by which Person B's salary exceeds Person A's.
(ii) Now, use the difference in natural logs to find the approximate percentage difference.

7 Suppose the following model describes the relationship between annual salary (*salary*) and the number of previous years of labor market experience (*exper*):

$$\log(salary) = 10.6 + .027\, exper.$$

(i) What is *salary* when $exper = 0$? When $exper = 5$? (*Hint*: You will need to exponentiate.)
(ii) Use equation (A.28) to approximate the percentage increase in *salary* when *exper* increases by five years.
(iii) Use the results of part (i) to compute the exact percentage difference in salary when $exper = 5$ and $exper = 0$. Comment on how this compares with the approximation in part (ii).

8 Let *grthemp* denote the proportionate growth in employment, at the county level, from 1990 to 1995, and let *salestax* denote the county sales tax rate, stated as a proportion. Interpret the intercept and slope in the equation

$$grthemp = .043 - .78\, sales\ tax.$$

9 Suppose the yield of a certain crop (in bushels per acre) is related to fertilizer amount (in pounds per acre) as

$$yield = 120 + .19\sqrt{fertilizer}.$$

(i) Graph this relationship by plugging in several values for *fertilizer*.
(ii) Describe how the shape of this relationship compares with a linear relationship between *yield* and *fertilizer*.

10 Suppose that in a particular state a standardized test is given to all graduating seniors. Let *score* denote a student's score on the test. Someone discovers that performance on the test is related to the size of the student's graduating high school class. The relationship is quadratic:

$$score = 45.6 + .082\, class - .000147\, class^2,$$

where *class* is the number of students in the graduating class.
(i) How do you literally interpret the value 45.6 in the equation? By itself, is it of much interest? Explain.
(ii) From the equation, what is the optimal size of the graduating class (the size that maximizes the test score)? (Round your answer to the nearest integer.) What is the highest achievable test score?
(iii) Sketch a graph that illustrates your solution in part (ii).
(iv) Does it seem likely that *score* and *class* would have a deterministic relationship? That is, is it realistic to think that once you know the size of a student's graduating class you know, with certainty, his or her test score? Explain.

11 Consider the line

$$y = \beta_0 + \beta_1 x.$$

(i) Let (x_1, y_1) and (x_2, y_2) be two points on the line. Show that $(\bar{x}, \bar{y})$ is also on the line, where $\bar{x} = (x_1 + x_2)/2$ is the average of the two values and $\bar{y} = (y_1 + y_2)/2$.
(ii) Extend the result of part (i) to n points on the line, $\{(x_i, y_i): i = 1, \ldots, n\}$.