Chapter S2 Review of Fundamentals

Contents

S2.1 Some Examples of Power Sets

Example S2.1 Find the power set of $\emptyset$.

Solution

The only possible subset of $\emptyset$ is itself and so $\mathcal{P}(\emptyset) = \{\emptyset\}$. Note that using the formula for finding the number of subsets of a set, we have $n = 0$ and $2^0 = 1$. ■

Example S2.2 Find the power set of $X = \{a\}$.

Solution

A set with only one element is usually called a **singleton;** it has two subsets, so $\mathcal{P}(X) = \{\{a\}, \emptyset\} = \{X, \emptyset\}$. ■

Example S2.3 Find the power set of $Z_+ = \{1, 2, 3, \ldots\}$.

Solution

We cannot construct the power set of Z_+ by enumeration, but we can define it by

$$\mathcal{P}(Z_+) = \{X : X \subseteq Z_+\}$$

the set of all possible sets of positive integers. ■

S2.2 Proof of Theorem 2.1

The proof is by contradiction. Suppose that $\sqrt{2} \in Q$. Then there exist integers p and q such that $p/q = \sqrt{2}$, where we choose the smallest such p and q. Now $p^2/q^2 = 2$, or $p^2 = 2q^2$, so p^2 must be an even number. Since the square of an odd number is always odd, p is even, and we may write $p = 2r$, where r is also an integer. We then have

$$p^2 = 4r^2 = 2q^2$$

and so $q^2 = 2r^2$. Clearly, q^2, and hence q, must be even. This result contradicts the assumption that p and q were the smallest numbers to give $p/q = \sqrt{2}$. Thus the statement $\sqrt{2} \in Q$ must be false. ■

This result establishes the fact that the line in figure 2.6 must contain at least one number that is not rational. Alternatively, it shows that if we wish to solve a simple equation such as

$$x^2 - 2 = 0$$

then we need a set of numbers larger than Q, since x cannot belong to Q. In fact there are a large number of irrational numbers.

S2.3 The Completeness Property of $\mathbb{R}$

We have already indicated that the set of real numbers $\mathbb{R}$ has the property that there are no "gaps." That is, between any two points (numbers) on the real line, every point is occupied by a number that is either a rational number or an irrational number. In other words, corresponding to each point on the real line there is a real number, and vice versa. This is known as the **completeness property,** and we can express this property formally by considering the concepts of the greatest lowest bound and the least upper bound of a subset of $\mathbb{R}$. We develop these ideas in the following definitions.

Definition S2.1

A set $S \subset \mathbb{R}$ is bounded above if there exists $b \in \mathbb{R}$ such that for all $x \in S$, $x \leq b$; b is then called an *upper bound* of S. A set T is bounded below if there exists $a \in \mathbb{R}$ such that for all $x \in T$, $x \geq a$; a is then called a *lower bound* of T.

For example, the set $Z_+ = \{1, 2, 3, \ldots\}$ is bounded below but unbounded above. The set $\bar{\mathbb{R}}_+ = \mathbb{R} - \mathbb{R}_+ = \{x \in \mathbb{R} : x < 0\}$ is unbounded below but bounded above.

If a subset of $\mathbb{R}$ has an upper (lower) bound, it has an infinity of upper (lower) bounds. This follows from the transitivity property of the inequality relation (part (ii) of theorem 2.2), since if $x \in S$ and $x \leq b$ and $b \leq y$, then $x \leq y$ and y is also an upper bound of S. However, we are interested in just one upper bound of S—the smallest upper bound, called the **supremum**—and just one lower bound of S—the largest lower bound, called the **infimum.**

Definition S2.2 The supremum of a set S, written sup S, has the properties:

(i) $x \leq \sup S$ for all $x \in S$.
(ii) If b is an upper bound of S, then $\sup S \leq b$.

Definition S2.3 The infimum of a set T, written inf T, has the properties:

(i) $x \geq \inf T$ for all $x \in T$.
(ii) If a is a lower bound of T, then $a \leq \inf T$.

Here we see that $\inf Z_+ = 1$ while $\sup Z_+$ does not exist; $\inf \bar{\mathbb{R}}_+$ does not exist while $\sup \bar{\mathbb{R}}_+ = 0$. These show that the inf and sup of a subset of $\mathbb{R}$ may or may not exist, and when either of these does exist, it may or may not be an element of the set. (Look at the sup of $\bar{\mathbb{R}}_+$.)

Theorem S2.1 If the sup or the inf of a subset of $\mathbb{R}$ exists, then it is unique.

The completeness property of $\mathbb{R}$ may now be stated as

Theorem S2.2 Every nonempty subset of $\mathbb{R}$ that has an upper bound has a supremum (least upper bound) in $\mathbb{R}$. Similarly every nonempty subset of $\mathbb{R}$ that has a lower bound has an infimum (greatest lower bound) in $\mathbb{R}$.

See the application of this theorem, for sup S, in figure S2.1.

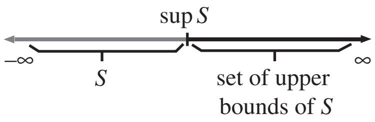

Figure S2.1 Completeness property of $\mathbb{R}$

S2.4 Proofs, The Necessary and Sufficient Conditions

Why should one person ever accept as true a statement made by someone else? The usual response would be, "Prove it!" If the statement is a purely factual one, for example, "prices have risen," then proof would take the form of some factual evidence that substantiates the statement. Economics is more often concerned, however, with deductive statements such as the *if . . . then* statement:

If the money supply increases, *then* the price level will rise.

That is to say, increases in the money supply lead to inflation. A stronger statement is the following:

The price level rises *if and only if* the money supply increases.

That is to say, *only* increases in the money supply lead to inflation. We are interested in how statements of this type are proved.

It is useful to express such statements in a general symbolic notation. We introduce the symbol "$\Rightarrow$" for the relation "if then," and the symbol "$\Leftrightarrow$" for "if and only if." We use capital letters such as P and Q to stand for basic statements such as "the money supply increases" (P) or "prices rise" (Q). We could then write the statements above as

$$P \Rightarrow Q$$

$$P \Leftrightarrow Q$$

There are several ways in which these statements can be read, and it is useful to spell these out:

$P \Rightarrow Q$ can be read

- if P then Q
- P implies Q
- P is a sufficient condition for Q
- Q is a necessary condition for P
- P only if Q

"P is sufficient for Q," means that the truth of P *guarantees* the truth of Q. Q is always true when P is true. It follows that if Q is not true, then P cannot be true. Thus it is *necessary* that Q is true for P to be true. In other words, P can be true *only if* Q is true.

We illustrate this in the Venn diagram of figure S2.2.

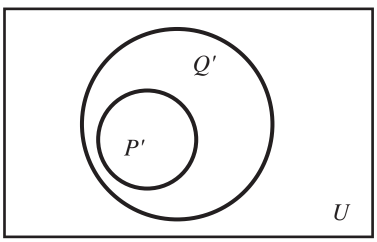

Figure S2.2 $P \Rightarrow Q$ means $P' \subseteq Q'$

The rectangle in figure S2.2 denotes the universal set of all possible cases. The set P' corresponds to the subset of cases for which the statement P is true and the set Q' the set of cases for which the statement Q is true. Then $P \Rightarrow Q$ can be interpreted as $P' \subseteq Q'$, since, whenever we have a case for which P is true, Q is also true, so P' must be contained in Q'. It follows that for a case to be in P', it must also be in Q', so it is *necessary* to be in Q' in order to be in P'.

$P \Leftrightarrow Q$ can be read

- P if and only if Q
- P is equivalent to Q
- P is a necessary and sufficient condition for Q
- P implies and is implied by Q

The last two statements reflect the fact that $P \Leftrightarrow Q$ means the same thing as "$P \Rightarrow Q$ and $Q \Rightarrow P$." So, in set theoretic terms, $P' \subseteq Q'$ and $Q' \subseteq P'$ so that $P' = Q'$ (see section 2.1).

The following economic example illustrates how we might prove a statement $P \Rightarrow Q$.

A Simple Model of the Quantity Theory of Money

We wish to prove the statement, "If the money supply increases, then the price level rises." Suppose that we have the following theory of the demand for money. The total amount of money that households in the economy want to hold is proportional to their level of nominal income. Nominal income is simply the product of real income (income measured at constant prices) Y and the price level P. So demand for money is kPY, where k is some positive constant. Now suppose that real income is fixed (e.g., because the economy is fully employing all of its resources) and the price level is free to vary. Finally, we assume that in equilibrium, money demand equals money supply M, or

$$M = kPY, \qquad k > 0$$

Money supply is taken to be determined exogenously by the government. With k and Y constant, we can write this equation simply as

$$P = aM, \qquad a \equiv \frac{1}{kY} \text{ (a constant)}$$

It is obvious from this that an increase in M, say from M_1 to M_2, leads to a change in price of

$$P_2 - P_1 = a(M_2 - M_1) > 0$$

Thus we have proved that if the money supply increases, the price level rises.

We want to illustrate two things with this example. First, it shows that proofs of propositions in economics must be model-specific. The proposition follows logically from the assumptions of our model. We have not proved that actual increases in the money supply lead to price rises—that is a matter for statistical and econometric analysis.

Second, the proof is an example of a direct proof. The following are examples of indirect proof.

Proof of the Contrapositive Proposition

The contrapositive statement to $P \Rightarrow Q$ is "not $Q \Rightarrow$ not P." In our example this would be:

If prices do not rise, then there is no increase in the money supply.

A contrapositive statement is equivalent to the original statement, or

$$(P \Rightarrow Q) \Leftrightarrow (\text{not } Q \Rightarrow \text{not } P)$$

and so proving the contrapositive is equivalent to proving the original proposition. To illustrate, our theory tells us that

$$P_2 - P_1 = a(M_2 - M_1)$$

The contrapositive is that $P_2 = P_1$, so we must have $M_1 = M_2$ since $a > 0$. ■

Proof by Contradiction or the "Reductio ad absurdum"

Let us *assume* that P is true but that Q is false (i.e., that Q does not follow from P), or $P \Rightarrow$ not Q. We need to show that this leads to a contradiction or a statement that is false. To illustrate with our example, we assume that $M_2 - M_1 > 0$ but $P_2 - P_1 \leq 0$. In our theory, these can only be true at the same time if $a \leq 0$, which is a contradiction to our assumption that $a > 0$. ■

PRACTICE EXERCISES

S2.1. The equilibrium price in a model of a simple market is found by setting supply equal to demand. Assuming that price can never be negative, show which one of the following propositions is true and which one is false:

(a) If the demand function in a market is

$$D = a - bp, \qquad a, b > 0$$

and the supply function is

$$S = \alpha + \beta p, \qquad \alpha, \beta > 0$$

then an equilibrium price exists.

(b) If the demand and supply functions are as in part (a) and $a > \alpha$, then an equilibrium price exists.

S2.2. Prove by contradiction that the empty set is a subset of every subset of some universal set.

S2.3. The overall effect of a change in the price of a good on the demand for it is the sum of two separate component effects: The substitution effect (demand for the good will increase when price falls because it is now cheaper relative to its substitutes), and the income effect (a fall in the price of a good increases the consumer's real income, leading to an increase in demand if the good is a *normal good* and a fall in demand if the good is an *inferior good*).

(a) Prove the following statements using each of the three methods of proof discussed in this section:

(i) A sufficient condition for the demand for a good to increase when its price falls is that it is a normal good.

(ii) A necessary but not sufficient condition for the demand for a good to decrease when its price falls is that it is an inferior good.

(b) In a Venn diagram, illustrate the relationships among the following four sets:

(i) The set of goods for which demand increases when prices fall

(ii) The set of goods for which demand falls when prices fall

(iii) The set of normal goods

(iv) The set of inferior goods

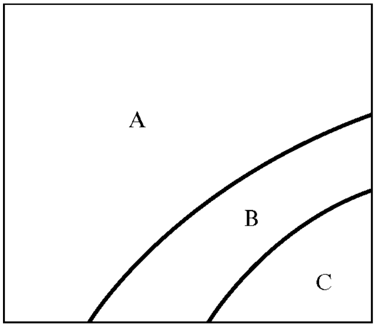

Figure S2.3 Venn diagram for exercise S2.3

Solutions

S2.1. Setting demand equal to supply, we obtain the equilibrium price $p^* = \frac{a-\alpha}{b+\beta}$. Since p^* cannot be negative, an equilibrium price will only exist when $a \geq \alpha$. Hence proposition 1 is false and proposition 2 is true.

S2.2. Let $S \subset U$, and $\bar{S}$ the complement of S. Assume $\{\} \not\subset S$. Then $\{\} \not\subset S \cap \bar{S}$; but then $\bar{S}$ cannot be the complement of S, since, by definition, $S \cap \bar{S} = \{\}$.

S2.3. **(a)** *Proof of statement (i):*

A sufficient condition for the demand for a good to increase when its price falls is that it is a normal good. Denoting A: "The good is normal" and B: "The demand for the good increases when its price falls," the statement can formally be written as A implies B. Now we are asked to prove it in the following ways: direct proof: A implies B; proof by contrapositive proposition: not B implies not A; and proof by contradiction: A implies not B leads to a contradiction.

- *Direct proof: A implies B.* When the price of the good falls, the consumer's real income increases. Since the good is normal by assumption, this signifies that the income effect is an increase in the demand for the good. The substitution effect has the same consequence: The fall in price leads to increased demand. Thus the total effect must be that the demand of the good rises.
- *Proof of the contrapositive proposition: not B implies not A.* If the demand for the good decreases when its price falls, it must be the income effect that causes decreased demand because the substitution effect always leads to increased demand. Since the fall in prices increases real income, the good must be inferior, that is, not normal.
- *Proof by contradiction: A implies not B leads to a contradiction.* A fall in the price of the good leads to an increase in real income. Because the good is normal by assumption, the income effect leads to an increase in demand for that good. If the demand for the good decreases when the price falls, this can only be due to the substitution effect. This contradicts the assumption that the substitution effect always leads to an increase in demand for a good whose price has fallen. Thus it must be true that when the good is normal, the demand rises when its price falls.

Proof of statement (ii):

A necessary but not sufficient condition for the demand for the good to decrease when its price falls is that it is an inferior good. Since the good is either normal or inferior, and its quantity demanded either falls or rises when its price falls (for simplicity we ignore the case where the demand stays the same), the second statement is simply: not B implies not A. But this was just proved.

(b) In the Venn diagram of figure S2.3, the rectangle shows the universal set, the set of all goods. Subset A is the set of normal goods, subset B the set of weakly inferior goods—the substitution effect is stronger than or equal to the income effect—and subset C is the set of strongly inferior goods—the income effect is stronger than the substitution effect. Then the answers are

(i) $A \cup B$

(ii) C

(iii) A

(iv) $B \cup C$

Chapter S3 Sequences, Series, and Limits

Contents

S3.1 More Applications of Series

Before moving on to applications, a general remark about series is in order. It may seem counterintuitive that a series such as the geometric series (with $0 < \rho < 1$), which is formed by adding up the first n terms of a sequence of which each term is a positive number, could have a limit (i.e., that the sum of an *infinite* number of terms, each with positive value, can have a finite value.) The following story, based on the classic problem of the **Achilles Paradox,** should convince you otherwise.

Example S3.1 The Achilles Paradox

Suppose that two families are taking identical cross-continental trips. Let us assume that the continent is a large one, like North America, with a stretch of 3,000 miles. Family A has early risers but relatively slow drivers. They leave at 7:00 a.m. and travel at 60 miles per hour (1 mile per minute). Family B, the next door neighbor, does not leave until 8:00 a.m. but travels at 120 miles per hour (2 miles per minute). Disregarding stops (including the possibility of a speeding ticket for family B), the astute reader will realize that family B will overtake family A after just one hour of travel (i.e., at 9:00 a.m.).

However, one can construct an argument to make it appear that family B *never* overtakes family A. We define an infinite sequence of time periods, all positive in value, during which family B reduces successive gaps that family A opens up without ever overtaking. To begin, family A has traveled to a point, call it P_1, that is 60 miles from its original location, the family home. It takes 30 minutes for family B to arrive at the same point P_1 (call this time interval τ_1 with $\tau_1 = 30$). In the meantime family A has traveled a further 30 miles to point P_2, and so it takes family B a further 15 minutes ($\tau_2 = 15$) to arrive at this point. During this time family A has, of course, traveled a further 15 miles to point P_3. Thus family B requires a further 7.5 minutes ($\tau_3 = 7.5$) to arrive at point P_3. As you can see, each time family B arrives at point P_i, a location family A previously arrived at, family A has time τ_i to open up another gap, which in turn takes family B time τ_{i+1} to "catch up." Since all periods τ_i are positive, family B does not catch up to family A even after an infinite number of (positive) intervals of time. Faulty intuition about the nature of an infinite series might then lead one to believe that family B will *never* catch up to family A, although common sense indicates that this is not so. The stages of this story are illustrated in figure S3.1 with D_A representing the distance covered by family A and D_B the distance covered by family B.

Hence we have an apparent paradox—the Achilles Paradox. This is because the paradox was first developed around 450 B.C. by the school of Greek philosophers called the Eleatics, of whom Zeno is the best-known member. In the original version of the story, the fleet-footed Achilles played the role of family B who left after family A but also drove more quickly. A nameless and slow-moving tortoise played the role of the early-rising but slow-driving family A.

The resolution of the Achilles Paradox resides in the fact that although one can construct an infinite sequence of positive time intervals, $\tau_i > 0, i = 1, 2, 3, \ldots,$ in which family B tries to catch up to family A but fails to do so, the sum of these time intervals forms an infinite series that converges:

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} \sum_{i=1}^{n} \tau_i$$

For the simple example above we know that the limit of this infinite series is 60 (minutes), which is the time it takes for family B to overtake family A. To see this formally, note that the series τ_i has terms $\tau_1 = 30$, $\tau_2 = 15$, $\tau_3 = 7.5, \ldots,$ or $\tau_i = a\rho^{i-1}$, with $a = 30$ and $\rho = 0.5$. Thus $s_n = \sum_{i=1}^{n} a_i$ is a geometric series and

$$\lim_{n\to\infty} s_n = \frac{a}{1-\rho} = 60$$

Having resolved the Achilles Paradox by recognizing that an infinite sum of positive numbers may be finite should not lead one to claim too much. In the geometric series, $a_n = a\rho^{n-1}$ with $|\rho| < 1$, the nth term approaches zero as $n \to \infty$. However, it is not generally correct that for *any* sequence a_n with $\lim_{n\to\infty} a_n = 0$, the associated series $s_n = \sum_{i=1}^{n} a_i$ converges. ■

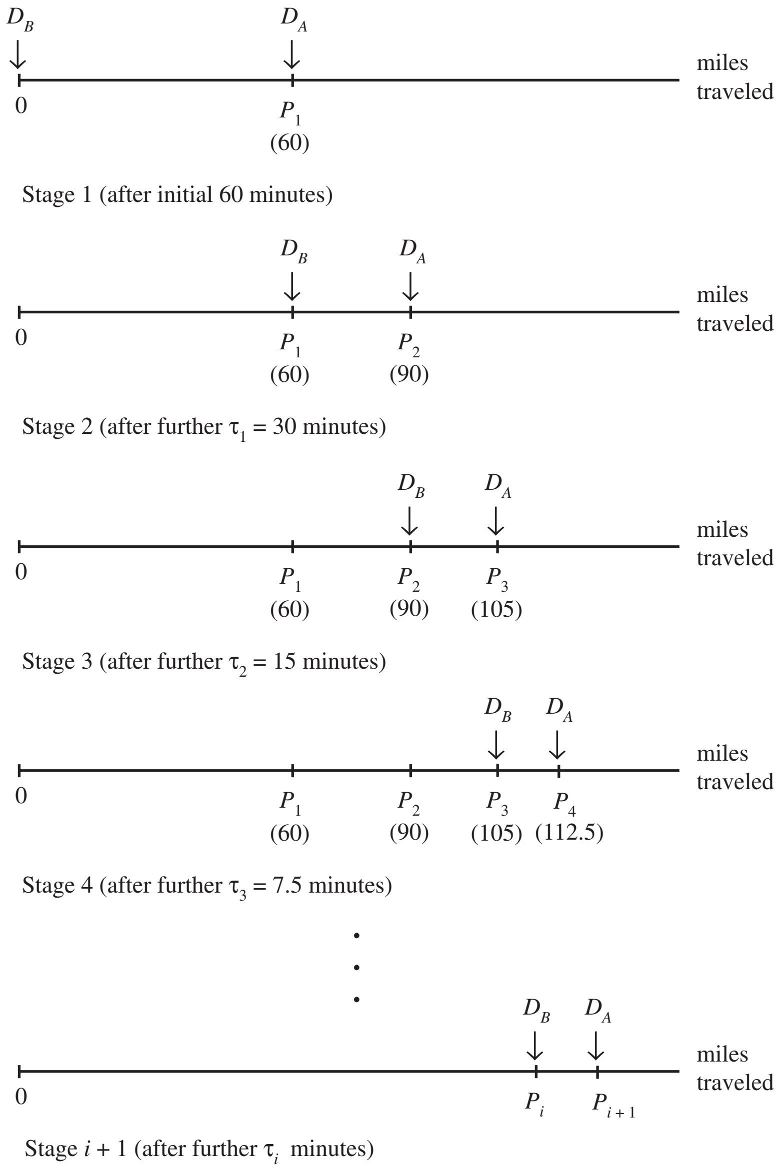

Figure S3.1 Relative locations of the families in the Achilles Paradox story

S3.2 The Keynesian Multiplier

An important component of the traditional Keynesian macroeconomic model used to explain the importance of government fiscal policy is the **multiplier.** The basic idea is that if there is an increase in exogenous expenditure in the economy, say government spending, then a multiplier effect ensues so that the ultimate impact on economic activity (GNP) is greater than the initial expenditure. For example, suppose that the government initiates additional expenditure of \$100 million for increased road repair. This expenditure then becomes additional income for firms and households. We would expect these individuals to save a certain fraction of this extra income, pay part of it as taxes, and perhaps use some to purchase imported goods, but we would also expect that a certain fraction would be spent on domestically produced goods and services. If the fraction of this income spent domestically is 60%, then the increased government expenditure will have a second-round impact on domestic incomes of \$60 million (i.e., \$100 million $\times$ 0.6). If the individuals receiving this additional income of \$60 million also spend a certain fraction domestically, and if this is also 60%, then the third-round effect is equal to \$36 million (i.e., \$60 million $\times$ 0.6). Continuation of this process is described in table S3.1:

Table S3.1

1st round effect (initial)	100.00 million	
2nd round effect	60.00 million	(100×0.6)
3rd round effect	36.00 million	(60×0.6)
4th round effect	21.60 million	(36×0.6)
5th round effect	12.96 million	(21.6×0.6)
and so on		

We can see in the table that the overall impact can be described as the geometric series

$$100 + 100(0.6) + 100(0.6)^2 + 100(0.6)^3 + \cdots$$

which, from equation (3.5), has the value ($a = 100$, $\rho = 0.6$)

$$\frac{100}{1 - 0.6} = \frac{100}{0.4} = 250 \text{ million}$$

Even if we allow for an infinite number of rounds in this problem, the overall increase in national income is finite because of the same intuition that resolves the Achilles Paradox. Summing an infinite number of terms, each positive, can give rise to a finite value if the terms become small sufficiently fast.

Example S3.2 Suppose that there is an initial increase in government spending of \$100 billion and that individuals spend 80% of any extra income on domestically produced goods. Use the Keynesian multiplier model to determine the overall impact of this \$100 billion injected into the economy. Also compute the first five rounds of the process as was done in table S3.1.

Solution

The first five rounds of effects are described in table 3.3. The overall impact is

$$100 + 100(0.8) + 100(0.8)^2 + 100(0.8)^3 + \cdots = \frac{100}{1 - 0.8} = 500 \text{ billion}$$

Table S3.2

1st round effect (initial)	100.00 billion	
2nd round effect	80.00 billion	(100×0.8)
3rd round effect	64.00 billion	(80×0.8)
4th round effect	51.20 billion	(64×0.8)
5th round effect	40.96 billion	(51.2×0.8)

■

In general, if we let c be the fraction of extra income that each individual spends on domestically produced goods ($0 < c < 1$), then the overall impact of an initial injection of amount E into the economy is

$$E + cE + c^2E + c^3E + \cdots = \frac{E}{1 - c}$$

Since $1 - c$ is the fraction of additional income *not* spent on domestically produced goods, we write $w = 1 - c$ and refer to it as the **propensity to withdraw** (i.e., money withdrawn from the cycle of spending on domestic goods and services). The fraction $1/w$ is known as the Keynesian multiplier.

S3.3 St. Petersburg Paradox

The expected monetary gain of a lottery, gamble, portfolio of stocks, or any other risky venture is computed by adding the probabilities of each possible outcome multiplied by the monetary value of the outcome. For example, consider a risk that involves winning nothing should a head be the outcome of a (fair) coin toss or winning \$20 million should a tail be the outcome. Since the probability of each possible outcome is 0.5, the expected value of the monetary gain is

$$0.5(0) + 0.5(\$20 \text{ million}) = \$10 \text{ million}$$

A natural question that arises is how to compare situations involving risk. For example, would one prefer the risky situation above to receiving \$5 million if a head appears and \$10 million if a tail appears (expected value = \$7.5 million)? A simple suggestion would be to postulate that individuals prefer those risks with the highest possible expected monetary gain. The example above might convince one that this postulate leaves something to be desired, as many individuals may well prefer the second lottery despite its lower expected monetary value. The Swiss mathematician Daniel Bernoulli offered the following type of gamble to illustrate even more dramatically the folly in adopting the expected monetary value rule as a behavioral postulate. The problem is now referred to as the St. Petersburg Paradox.

Consider the gamble involving the successive tossing of a (fair) coin with the player receiving a prize of $\$2^n$ when heads occurs for the first time after n tosses (e.g., if the first three coin tosses are tails and the fourth is a head, the player wins $\$2^4$ or \$16). Since coin tosses are independent events, the probability that the first time a head appears on the nth toss is $(1/2)^n$ (e.g., the probability of three successive tails followed by a head is $(1/2)^4 = 1/16$). The gamble could, in principle, involve an arbitrarily large number of coin tosses. The expected monetary gain of the gamble is

$$EM = \sum_{n=1}^{\infty} \left(\frac{1}{2^n}\right) 2^n = 1 + 1 + 1 + \cdots = \infty \tag{S3.1}$$

which is clearly a divergent series. This means that if one is willing to accept the idea that individuals value risky outcomes according to the expected value of the monetary outcome, then this gamble is preferred to a gain of \$10 million for certain or any other finite amount, no matter how large. Bernoulli believed that this was a ludicrous conclusion and offered a way out of the apparent paradox.

An intuitive explanation of his argument is as follows. The value to an individual of different monetary awards is not simply measurable in dollars, since,

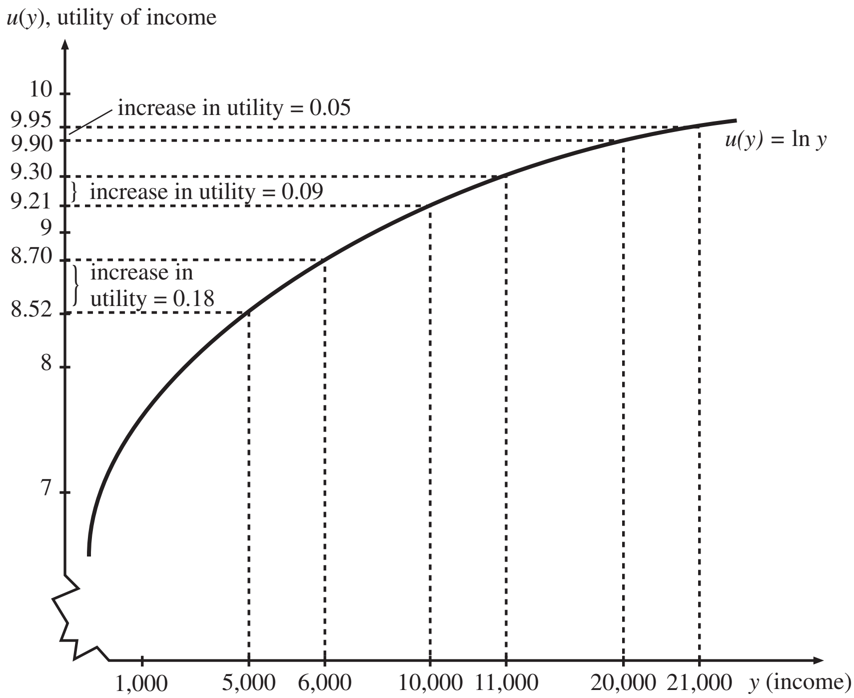

Figure S3.2 Natural log function for utility of income. Note that the increase in utility resulting from an extra \$1,000 is less the greater is the initial income.

for example, the *value* of the first \$1 million received is not likely equivalent to the value of a second \$1 million. Rather, one should assign *utility values* to the monetary outcomes with the increase in utility for a given increase in income being less the greater the initial income. Intuitively this means that the first million dollars is *worth more* to an individual than the second million dollars. In particular, Bernoulli suggested using the natural logarithm function to generate utility values so that the utility of $\$y$ is $\ln(y)$. Using this function, one can see that the increase in utility resulting from an extra dollar of income gets smaller as income increases (see figure S3.2). If one compares risky situations by using the expected value of utilities rather than money, the expected utility of the St. Petersburg Paradox game is

$$EU = \sum_{n=1}^{\infty} \left(\frac{1}{2^n}\right) \ln(2^n) \tag{S3.2}$$

which turns out to be a series that converges (i.e., has a finite-valued limit). We can see this immediately by using theorem 3.4. That is,

$$\begin{aligned}\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| &= \lim_{n\to\infty}\frac{\left[1/2^{(n+1)}\right]\ln\left(2^{(n+1)}\right)}{\left[1/2^n\right]\ln(2^n)}\\ &= \lim_{n\to\infty}\frac{2^n}{2^{(n+1)}}\frac{(n+1)\ln 2}{n\ln 2}\\ &= \lim_{n\to\infty}\frac{(n+1)}{2n}\\ &= \lim_{n\to\infty}\left(\frac{1}{2}+\frac{1}{n}\right)=\frac{1}{2}\end{aligned}$$

which is less than 1, and so the series represented by equation (S3.2) converges. (Note that the second line follows from $\ln(2^{n+1}) = (n+1)\ln 2$ and $\ln(2^n) = n\ln 2$.) Thus, by using the function $\ln(y)$ to represent the utility of $\$y$ and presuming that individuals rank risky alternatives by comparing the expected value of the utility they generate, rather than the expected monetary value, we can evade the problem of concluding that a St. Petersburg gamble is worth more than any certain amount of money, no matter how large in value.

Examining this issue further is worthwhile as it brings out some of the intuition about infinite series of terms that tend to zero or infinity. Let p_n be the sequence that represents the probability that the first head occurs after n tosses (i.e., $p_n = 1/2^n$), and let $u_n = u(y_n)$ be the utility gain upon receiving the winnings of y_n (i.e., $y_n = 2^n$ and $u(\cdot) = \ln(\cdot)$ for the example above. Thus

$$EU = \lim_{N\to\infty}\sum_{n=1}^{N} p_n u_n$$

represents the expected utility of the gamble. The limit is the sum of the product of the terms from two sequences, one that tends to zero (p_n) and another that tends to infinity (u_n) as $n \to \infty$. The series converges if the u_n terms do not approach infinity *too quickly* relative to how quickly the p_n terms approach zero. We saw above that if $y_n = 2^n$ and we use $u_n = y_n$ (or any linear, increasing function of y_n), then the series does not converge, while if we use $u_n = \ln(y_n)$, the series does converge. The problem with this resolution of the St. Petersburg Paradox is that one can construct a sequence of prizes, such as $y_n = e^{2^n}$, such that the series will no longer converge even if $u(\cdot) = \ln(\cdot)$. This follows since

$$
\begin{aligned}
EU &= \lim_{N\to\infty} \sum_{n=1}^{N} p_n u_n \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \left(\frac{1}{2^n}\right) \ln\left(e^{2^n}\right) \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \left(\frac{1}{2^n}\right) 2^n \ln(e) = 1 + 1 + 1 + \cdots
\end{aligned}
$$

which is a divergent series (recall that $\ln(e) = 1$). Thus, by altering the prize structure, we see that adopting the utility function $u(y) = \ln(y)$ will not always resolve the type of paradox raised by the St. Petersburg gamble.

A sufficient condition for resolving this paradox is to assume that the utility function is bounded. If $u(y)$ is bounded above by the value $u^{\max}$, as illustrated in figure S3.3, then the series

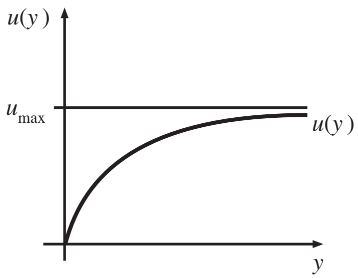

Figure S3.3 A bounded utility function

$$
\begin{aligned}
EU &= \lim_{N\to\infty} \sum_{n=1}^{N} p_n u_n \\
&< \lim_{N\to\infty} \sum_{n=1}^{N} p_n u^{\max} \\
&= \left(u^{\max}\right)\left(\lim_{N\to\infty} \sum_{n=1}^{N} p_n\right) \\
&= u^{\max}
\end{aligned}
$$

An alternative way out of the St. Petersburg Paradox is to assume that the prize values are bounded, say because of a budget constraint.

PRACTICE EXERCISES

S3.1. Suppose that there is an initial increase in government spending of \$20 billion and individuals spend 70% of any extra income on domestically produced goods. Find the overall impact on the economy according to the Keynesian multiplier model.

S3.2. *Consider a St. Petersburg gamble with the following information (see the notation in section 3.5): $p_n = 1/2^n$, $y_n = 2^n$, and $u(y) = y^{1/2}$.

(a) Use theorem 3.4 to show that EU, expected utility, is bounded.

(b) Find the level of expected utility for this case. Show that a consumer would prefer the certainty of \$2.50 to this gamble.

(c) Construct a sequence of prizes y_n so that EU is not bounded.

S3.3. Suppose that there is an initial increase in government spending of \$50 billion and individuals spend 75% of any extra income on domestically produced goods. Find the overall impact on the economy according to the Keynesian multiplier model.

S3.4. Consider a St. Petersburg gamble with the following information: $p_n = 1/2^n$, $y_n = 2^n$, $u(y) = y^2$. Show that the expected utility of this gamble is unbounded.

Solutions

S3.1. The first five round of effects are

1st round effect	20.00 billion	
2nd round effect	14.00 billion	(20×0.7)
3rd round effect	9.80 billion	(14×0.7)
4th round effect	6.86 billion	(9.8×0.7)
5th round effect	4.802 billion	(6.86×0.7)

The overall impact is

$$\begin{aligned} 20 + 20(0.7) + 20(0.7)^2 + 20(0.7)^3 + \cdots &= \frac{20}{1 - 0.7} \\ &= 66.67 \text{ billion} \end{aligned}$$

S3.2. **(a)**

$$\begin{aligned} EU &= \lim_{N\to\infty} \sum_{n=1}^{N} p_n u_n \\ &= \lim_{N\to\infty} \sum_{n=1}^{N} \frac{1}{2^n} (y_n)^{1/2} \end{aligned}$$

$$= \lim_{N\to\infty} \sum_{n=1}^{N} \frac{(2^n)^{1/2}}{2^n}$$
$$= \lim_{N\to\infty} \sum_{n=1}^{N} \frac{1}{(2^n)^{1/2}}$$
$$= \lim_{N\to\infty} \sum_{n=1}^{N} \frac{1}{2^{n/2}}$$

So EU is a series formed from the sequence $a_n = \frac{1}{2^{n/2}}$. Therefore we can construct the ratio

$$\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{\frac{1}{2^{(n+1)/2}}}{\frac{1}{2^{n/2}}}\right| = \frac{2^{n/2}}{2^{n/2+1/2}} = \frac{1}{2^{1/2}} < 1$$

and so, according to theorem 3.4, this series converges.

(b) Upon expanding the expression for EU, we get

$$EU = \frac{1}{2^{1/2}} + \frac{1}{2} + \frac{1}{2^{3/2}} + \frac{1}{2^2} + \cdots$$

We can easily see that this is a geometric series with first term $a = \frac{1}{2^{1/2}}$ and ratio $\rho = \frac{1}{2^{1/2}}$, and so we can use the formula

$$\frac{a}{1-\rho} = \frac{\frac{1}{2^{1/2}}}{1 - \frac{1}{2^{1/2}}} = \frac{1}{2^{1/2} - 1} \doteq 2.44$$

to determine that the value of expected utility of this gamble is approximately 2.44. Therefore a consumer would prefer to receive with certainty the amount \$6.25, for example, rather than the stated gamble (since $u(6.25) = 6.25^{1/2} = 2.5$).

(c) If we used the sequence of prizes $y_n = (2^n)^2 = 2^{2n}$, for example, then EU of this gamble would not be bounded. To see this is so, note that in this case

$$EU = \lim_{N\to\infty} \sum_{n=1}^{N} \frac{(2^{2n})^{1/2}}{2^n}$$

$$
\begin{aligned}
&= \lim_{N\to\infty} \sum_{n=1}^{N} \frac{2^n}{2^n} \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} 1 \\
&= \lim_{N\to\infty} N = +\infty
\end{aligned}
$$

S3.3. Overall impact is

$$
\begin{aligned}
&50 + 50(0.75) + 50(0.75)^2 + 50(0.75)^3 + \cdots \\
&= \frac{50}{1-0.75} = \frac{50}{0.25} = \$200 \text{ billion}
\end{aligned}
$$

S3.4. The expected utility of this gamble is

$$
EU = \lim_{N\to\infty} \sum_{n=1}^{N} p_n u_n
$$

where $p_n = \frac{1}{2^n}$ and $u_n = y_n^2 = (2^n)^2 = 2^{2n}$. Therefore we get

$$
\begin{aligned}
EU &= \lim_{N\to\infty} \sum_{n=1}^{N} \frac{1}{2^n} 2^{2n} \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} 2^n = +\infty
\end{aligned}
$$

Chapter S4 Continuity of Functions

Contents

S4.1 Revenue Function, Cost Function, and Profit Function for a Perfectly Competitive Firm

In the model of perfect competition it is assumed that each firm treats the market price as given. The firm does not believe its own choice of output level will influence the market price, and so it treats this value as fixed. This assumption is usually only made to describe markets in which a large number of producers each produces a small amount of some homogeneous (identical) product. Letting $\bar{p}$ be the price of the product and y be the firm's output level, we get the revenue function $R(y) = \bar{p}y$. Since this is a linear function in y, it is continuous. To show what must be true in order that the cost function, $C(y)$, be continuous takes a little effort. We begin with the long-run situation.

The total cost of producing a given output level is simply the cost of all the inputs used in the production process. Suppose that there is only one input x used to produce y and that the production function $y = f(x)$ is used to describe the production relation. If we know how much output can be created from various levels of input, then we can work backward to determine the level of input needed to produce a given level of output. This simply generates the inverse function

$$y = f(x) \Rightarrow x = f^{-1}(y)$$

as demonstrated in figure S4.1 for the case $y = x^{1/2}$.

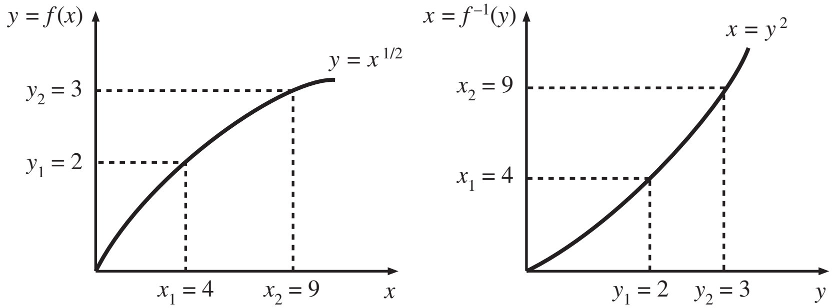

Figure S4.1 A production function and its inverse

Thus the cost of producing output level y is wx, where x is the amount of input required to produce output level y and w is the per-unit price of that input. Using the inverse function $x = f^{-1}(y)$, we can write $C(y) = wf^{-1}(y)$. Now, if the production function $f(x)$ is continuous, then, because of theorem 4.1(vi), so is its inverse, $f^{-1}(y)$, and so $C(y)$ is continuous. The profit function, $\pi(y) = R(y) - C(y)$ is also continuous by theorem 4.1(iii). This is illustrated in figure S4.2 for a more general cost function.

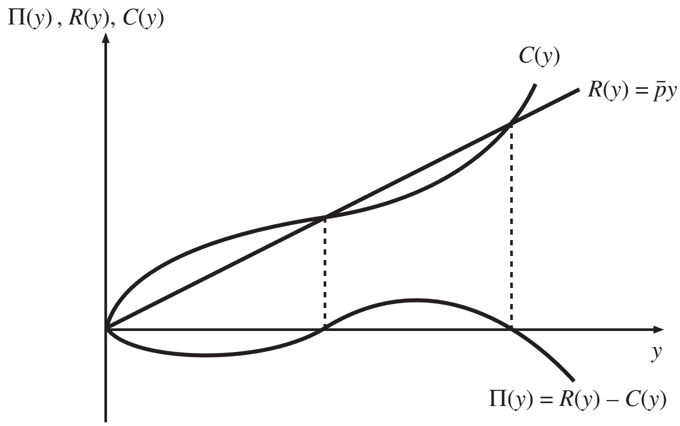

Figure S4.2 If the cost function, $C(y)$, is continuous, so is the profit function, $\pi(y)$.

In the short-run situation we usually model the firm's decision in terms of a single (variable) input. Again, let x represent this input and $y = f(x)$ the (short-run) production function. The same analysis as in the above paragraph applies except that we need to recognize that if the firm produces no output in the short

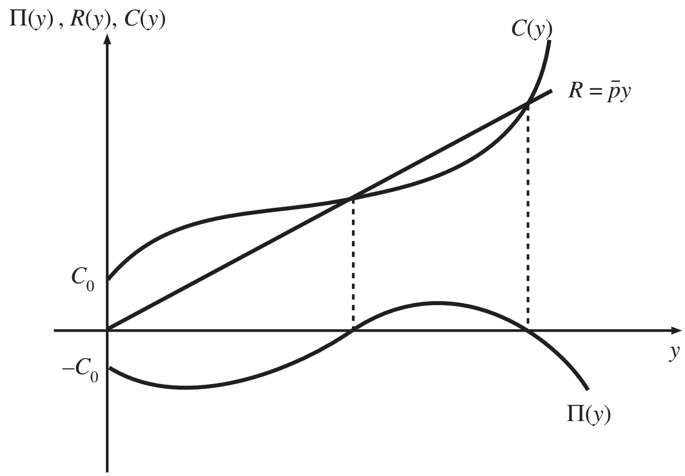

Figure S4.3 A cost function and profit function in the presence of fixed costs

run, it still must pay its fixed costs. If we let fixed cost be \$$C_0$, then the cost function becomes $C(y) = wf^{-1}(y) + C_0$. Once again it follows that if $f(x)$ is a continuous function, then so is the cost function and then so is the profit function. An example is illustrated in figure S4.3.

Consider the following case in which the profit function is not continuous: Suppose that a firm cannot change from zero production to positive production without expending resources to start the production process. For example, a smelting furnace may have to be preheated before any steel can be produced. This type of cost is called a *setup cost* and differs from the usual sort of cost in that it is a lump-sum cost that must be incurred when going from zero to any positive amount of production. The amount produced does not affect the size of this cost. The result is that the cost function, and hence the profit function, will be discontinuous at $y = 0$. Letting $c(y)$ represent the cost of producing y, *excluding* the setup costs, we can write the cost function as

$$C(y) = \begin{cases} 0, & y = 0 \\ B_0 + c(y), & y > 0 \end{cases}$$

where B_0 is the setup cost. Then the profit function is

$$\pi(y) = \begin{cases} 0, & y = 0 \\ R(y) - B_0 - c(y), & y > 0 \end{cases}$$

These functions are illustrated in figure S4.4.

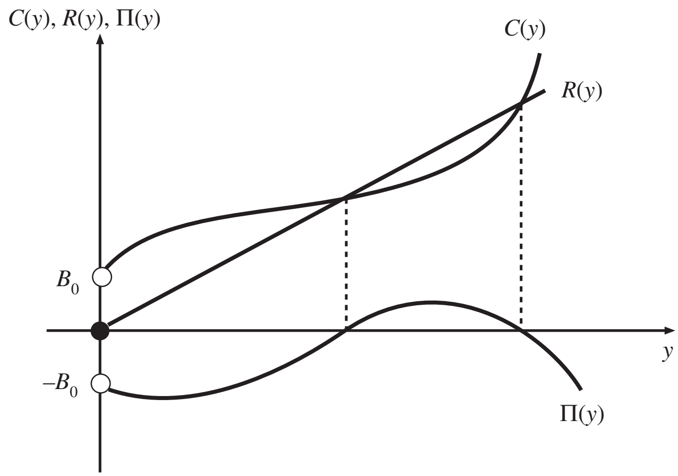

Figure S4.4 A cost function and profit function in the presence of setup costs

In comparing the cost function in figure S4.3 with that in figure S4.4, it is important to distinguish between fixed costs, a short-run phenomenon only, and setup costs. The existence of fixed costs does not lead to a discontinuity of the short-run cost function because, if the firm chooses to produce zero output, it is still the case that the firm must pay its fixed costs. Thus, for the cost function represented in figure 4.14, the right-hand limit of $C(y)$ at $y = 0$ is equal to C_0; that is to say, $\lim_{y \to 0^+} C(y) = C_0$, which is also equal to the value of the cost function at $y = 0$. For the cost function illustrated in figure S4.4, the setup costs, B_0, are incurred for any level of output no matter how small or large but are avoided when the firm chooses to produce zero output. Thus, the right-hand limit of $C(y)$ at $y = 0$ for this cost function is equal to B_0, namely $\lim_{y \to 0^+} C(y) = B_0$, but the value of the cost function at $y = 0$ is 0. Clearly, this function is discontinuous at the point $y = 0$.

S4.2 Hotelling's Location Model

Hotelling's location model is designed to illustrate why a group of firms selling the same product sometimes cluster together geographically when it seems that consumers would be better served if the firms were at different locations. For example, we often see more than one gas station or convenience store located across the street from each other or next to each other. Hotelling's location model explains how such a pattern can be the result of rational profit-maximizing behavior by firms. What is interesting about this model from the perspective of this chapter is that the key to the equilibrium solution of this model is, as in the Bertrand oligopoly model, the result of a discontinuity.

We use a highly stylized set of assumptions to model this phenomenon. Suppose that there are two firms, A and B, who will locate on a street which is represented by a straight line of length one mile. We can indicate any location on this street by a number, L, belonging to the closed interval [0, 1]. We assume that the product each firm sells is identical. The firms charge the same price $\bar{p}$ and face the same unit cost of production that is the constant value $\bar{c}$ with $\bar{p} - \bar{c} > 0$. Thus the profit on each unit is $\bar{p} - \bar{c}$ and is positive. Since the product the firms sell is identical (homogeneous good) and the price is the same, we assume that consumers will go to the firm that is closest and if the firms locate at the same point, one-half of the consumers will go to each firm. We also assume that the consumers are spread uniformly along the street and that each consumer buys one unit of the product each period. If there are N consumers, then the potential aggregate profit in the market is $N(\bar{p} - \bar{c})$. The strategic problem for each firm is then to choose a location in such a way as to maximize market share, since the greater is a firm's market share, the greater is its profit. The decision for each firm will depend on where the other firm is located.

Figure S4.5 Firm B's location at $L = 0.2$

To solve this model, we first treat firm B's location, $L_B \in [0, 1]$, as already determined at $L_B = \bar{L}_B$, and then see how firm A's market share changes as a function of its location, L_A, along the same interval [0, 1]. By considering where firm A will locate relative to $\bar{L}_B$, we can then consider where firm B should locate to obtain its greatest possible market share. In this series of steps we let the specific location for firm B be $\bar{L}_B = 0.2$, as illustrated in figure S4.5.

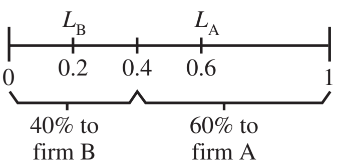

Figure S4.6 Market shares if firm B locates at 0.2 and firm A locates at 0.6

We use figure S4.5 to determine the market share of each firm. Say that firm B locates at $\bar{L}_B = 0.2$ and firm A locates at $L_A = 0.6$, as illustrated in figure S4.6. The midpoint between 0.2 and 0.6 is 0.4. Thus all consumers to the left of point 0.4 (40% of the market) will make their purchases from firm B while all those to the right of point 0.4 (60% of the market) will make their purchases from firm A.

To solve this model, it is easiest to determine what happens to firm A's market share, $M^A(L_A)$, as it changes its location decision beginning from 0 and moving toward 1 ($L_A \in [0, 1]$). If firm A locates at point $L_A = 0$, then firm A's market share will be 10% and firm B's will be 90%. This follows from the fact that the midpoint between 0 and 0.2 is 0.1. All consumers to the right of point 0.1 (90% of the market) will make their purchases from firm B, while those to the left of 0.1 (10% of the market) will make their purchases from firm A. Now, as firm A moves to the right, but still chooses $L_A < 0.2$, the midpoint between L_A and $\bar{L}_B = 0$ will also shift to the right, so more of the consumers will make their purchases from firm A. As long as $L_A < 0.2$, firm A will increase its market share gradually for small movements to the right. As firm A's location approaches $\bar{L}_B = 0.2$ from the left, its market share rises steadily and continuously to 20%. However, once firm

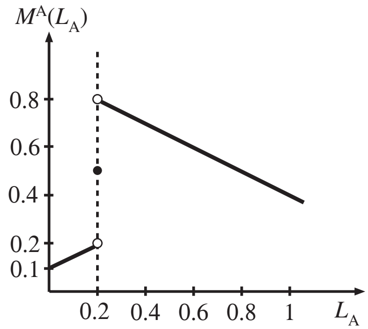

Figure S4.7 Market share of firm A as a function of its location, L_A, given B's location at $L_B = 0.2$

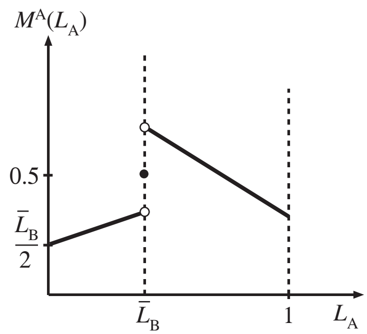

Figure S4.8 Market share for firm A as a function of its location, L_A, given a choice of location by B of $\bar{L}_B < 0.5$

A locates exactly at the point $\bar{L}_B = 0.2$, the two firms share the market equally as consumers are indifferent between them. Thus, at the point $L_A = 0.2$, firm A's market share *jumps* discontinously to 50%. If firm A locates a *small* distance, $\epsilon > 0$, to the right of $\bar{L}_B = 0.2$, its market share jumps again to *almost* 80%, since all consumers to the right of $0.2 + \epsilon$ make their purchases from firm A with the others going to firm B. As we increase the distance that firm A moves to the right of $\bar{L}_B = 0.2$, we find that A loses some of its market share as those consumers to the left of the midpoint between $\bar{L}_B$ and L_A ($\bar{L}_B < L_A$) will make their purchases from firm B. For example, if firm A locates at the rightmost point, $L_A = 1$, all the consumers to the left of 0.6 travel to firm B (60% of the market) while all the consumers to the right of 0.6 (40% of the market) travel to firm A.

From this information we can determine the market share for firm A as a function of location L_A for the given location choice of firm B, $\bar{L}_B = 0.2$. This is given in equation (S4.1) and is illustrated in figure S4.7:

$$M^A(L_A) = \begin{cases} L_A + 0.5(0.2 - L_A), & L_A < 0.2 \\ 0.5, & L_A = 0.2 \\ (1 - L_A) + 0.5(L_A - 0.2), & L_A > 0.2 \end{cases} \tag{S4.1}$$

We do the same for the general choice $L_B = \bar{L}_B$. The market-share function is given in equation (S4.2) and illustrated in figure S4.8:

$$M^A(L_A) = \begin{cases} L_A + 0.5(\bar{L}_B - L_A), & L_A < \bar{L}_B \\ 0.5, & L_A = \bar{L}_B \\ (1 - L_A) + 0.5(L_A - \bar{L}_B), & L_A > \bar{L}_B \end{cases} \tag{S4.2}$$

Looking at the market-share function $M^A(L_A)$ and returning to the conditions for continuity given in definition 4.3, we can see formally for the case of $\bar{L}_B = 0.2$ why this function is discontinuous at the point $L_A = 0.2$. At point $L_A = 0.2$ the left-hand limit of the market-share function is 0.2, the right-hand limit is 0.8, and the value of the function itself is 0.5. Thus we have

$$\lim_{L_A \to 0.2^-} M^A(L_A) = 0.2, \quad \lim_{L_A \to 0.2^+} M^A(L_A) = 0.8, \quad M^A(0.2) = 0.5$$

Since these values are all different, the function is not continuous at the point $L_A = 0.2$. In fact only two of these values need differ for the function $M^A(L_A)$ to be discontinuous.

From an economic standpoint this discontinuity is extremely important in the model, since what happens at the point of discontinuity drives the model to its solution. To see why this is so, suppose that firm B happens to locate to the left of center (i.e., $\bar{L}_B < 0.5$). In this case firm A's market-share equation slopes upward from $L_A = 0$ to the point where $L_A = \bar{L}_B$. If located just to the left of $L_A = \bar{L}_B$, firm A's market share equals the value $\bar{L}_B < 0.5$ and so is less than 50%. When located at precisely the same point as firm B (e.g., the other side of the street but the same distance to each end), each firm acquires 50% of the market. However, if firm A locates just to the right of $\bar{L}_B$, then firm A will be servicing all consumers to the right of $\bar{L}_B$, which means its market share will be greater than 50% (since $1 - \bar{L}_B > 0.5$). To move any further right will mean that firm A starts to lose market share, and so the best thing for firm A to do is to locate just to the right of $\bar{L}_B$ and be rewarded with more than 50% of the market. A similar argument can be made to show that if firm B locates to the right of center ($\bar{L}_B > 0.5$), then the best thing for firm A to do is to locate just to the left of firm B and once again be rewarded with more than 50% of the market. Thus, if firm B locates on one or the other side of center, the result will be that firm B gets less than 50% of the market and firm A gets more than 50% of the market. Since firm B can presumably determine that this would be the outcome, then firm B can get the greatest market share for itself by locating just at the center. In this case the best firm A can do is locate just on either side of center (or also at the center), with the result that both firms acquire 50% of the market, as illustrated by figure S4.9. The same argument would apply in reverse if we considered the scenario in which firm A is treated as if it is the firm that locates first. The outcome that both firms locate at the center is an equilibrium for the model. This is because if both locate at the center, neither can do better by altering its decision. Thus, by careful consideration of the discontinuity in the market-share function, we are able to solve this model.

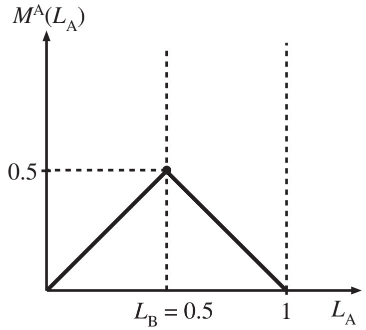

Figure S4.9 Market share of firm A as a function of its location, L_A, given the choice of location $\bar{L}_B = 0.5$ by firm B

S4.3 Intermediate-Value Theorem

In this section we present a straightforward theorem, the intermediate-value theorem and show that it can be very powerful in the study of equilibrium, which is one of the most important concepts in economics. The particular application we make is a very simple one. The range of applications, however, is in fact very broad.

Suppose that the function $y = f(x)$ is continuous on the interval $[a, b]$, $b > a$. It follows that the function must take on every value between $f(a)$ and $f(b)$, which are the function values at the endpoints of the interval $[a, b]$. This result is called the intermediate-value theorem because any intermediate value between $f(a)$ and $f(b)$ must occur for this function for at least one value of x between $x = a$ and $x = b$. This result is understood intuitively by looking at figures S4.10 and S4.11. In figure S4.10 it is clear that any value between $y = f(a)$ and $y = f(b)$, for example, $y = \bar{y}$, is realized by the continuous function $f(x)$ for some $x \in [a, b]$. In

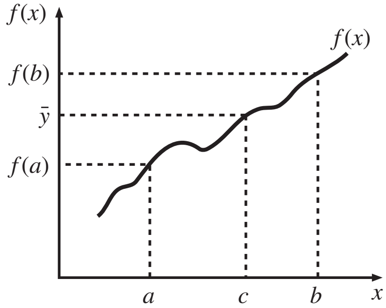

Figure S4.10 Demonstration of the intermediate-value theorem for a function $f(x)$ that is continuous on $[a, b]$

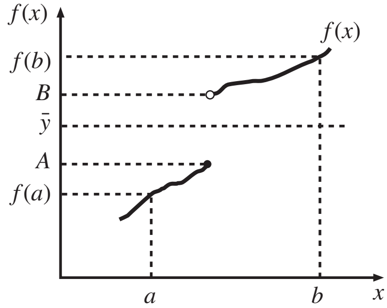

Figure S4.11 Demonstration that if $f(x)$ is discontinuous on $[a, b]$, then it doesn't necessarily take on every value between $f(a)$ and $f(b)$

figure S4.11 we see that this is not necessarily the case for a discontinuous function as no value between $y = A$ and $y = B$ is realized by the function for any $x \in [a, b]$.

Theorem S4.1 **(Intermediate-value theorem)** Suppose that $f(x)$ is a continuous function on the closed interval $[a, b]$ and that $f(a) \neq f(b)$. Then, for any number $\bar{y}$ between $f(a)$ and $f(b)$, there is some value of x, say $x = c$, between a and b such that $\bar{y} = f(c)$.

The Existence of Equilibrium

The simple result of the intermediate-value theorem is often very useful when trying to prove that a special value of an economic variable exists. Consider the simple partial equilibrium model of demand and supply with p representing price, y representing quantity, $y = D(p)$ representing the demand function, and $y = S(p)$ representing the supply function. An equilibrium price for this model is defined as a price that clears the market. That is, $p = p^e \geq 0$ is an equilibrium price if $D(p^e) = S(p^e)$. We sometimes include the possibility of a *free good* which is formally defined as the case where $p^e = 0$ if $D(0) < S(0)$. It is clear that if the demand curve intersects the supply curve at a point (y^e, p^e) with both $y^e > 0$ and $p^e > 0$, as in figure S4.12, then this is an equilibrium. The case of an equilibrium for a free good is given in figure S4.13.

Rather than work with the two schedules, demand $D(p)$ and supply $S(p)$, it is useful to amalgamate them into a single function called the excess-demand function, $z(p) = D(p) - S(p)$. Notice that the value of $z(p)$ indicates the amount by which demand exceeds supply if $z(p) > 0$, while if $z(p) < 0$, the absolute

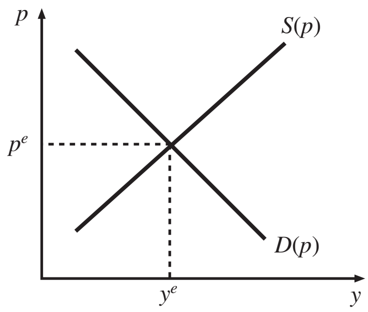

Figure S4.12 Market equilibrium with positive equilibrium price and output

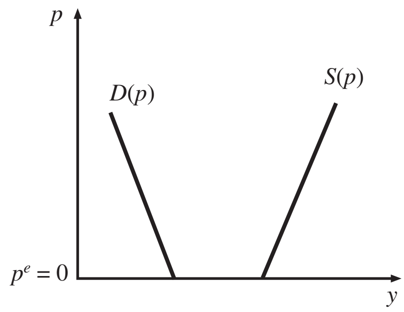

Figure S4.13 Market equilibrium for a free good

value of $z(p)$ indicates the amount by which supply exceeds demand. In terms of the excess-demand function, $z(p^e) = 0$ whenever $D(p^e) = S(p^e)$, while $z(0) \leq 0$ if $D(0) \leq S(0)$. Thus an alternative description of an equilibrium price is that $p^e > 0$ is an equilibrium price provided $z(p^e) = 0$, while $p^e = 0$ is an equilibrium price provided $z(0) \leq 0$. For the case in figure S4.12, excess demand is positive whenever $p < p^e$, while excess demand is negative whenever $p > p^e$. The excess-demand functions for each of the cases in figures S4.12 and S4.13 are given in figures S4.14 and S4.15 respectively.

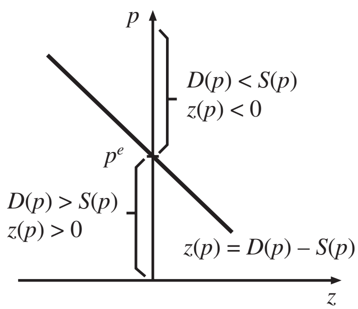

Figure S4.14 Excess-demand function for the case of market equilibrium shown in figure S4.12

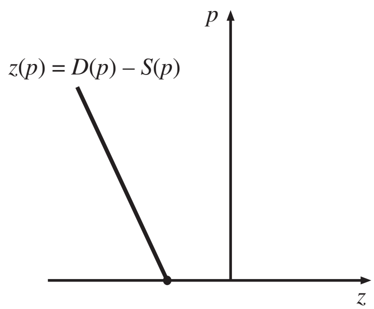

Figure S4.15 Excess-demand function for the case of market equilibrium shown in figure S4.13

Notice that in the figures, the variable p is placed on the vertical axis and the variables that are functions of p—$D(p)$, $S(p)$ and $z(p)$—are placed on the horizontal axis. This choice is not the conventional one in mathematics but is generally used in economics when dealing with demand- and supply-functions.

It is useful to have a set of sufficient conditions that guarantee that an equilibrium price exists. The intermediate-value theorem (theorem S4.1) helps us to do just that. Suppose that we consider a commodity that requires costly resources to produce and so at a price of zero supply would be zero, $S(0) = 0$. Further assume that $D(0) > 0$; that is, that the good is desirable from the consumers' point of view. Then, at $p = 0$, we have

$$z(0) = D(0) - S(0) > 0$$

Moreover let us believe that at least for prices above some *sufficiently* high price, call this level $p = \hat{p}$, firms will find it so profitable to produce this product and consumers will find the price so high that supply will exceed demand. Thus $D(\hat{p}) < S(\hat{p})$ and so $z(\hat{p}) = D(\hat{p}) - S(\hat{p}) < 0$. Now, if the demand and supply functions are continuous on the interval of prices $p \in [0, \hat{p}]$, then so will $z(p)$ be continuous on $p \in [0, \hat{p}]$, according to theorem 4.1(iii). Thus, by the intermediate-value theorem, every function value between $z(0)$ and $z(\hat{p})$ must be realized by some value p between 0 and $\hat{p}$. In particular, since $z(0) > 0$ and $z(\hat{p}) < 0$, there must be some value of p between 0 and $\hat{p}$, say $p = c$, such that $z(c) = 0$. But such a value of p is by definition an equilibrium price (i.e., $c = p^e$). Theorem S4.2 provides sufficient conditions under which we are guaranteed that a positive equilibrium price exists.

Theorem S4.2 If the demand and supply functions are continuous and the following two conditions are satisfied:

(i) at zero price, demand exceeds supply, $D(0) > S(0)$, meaning that $z(0) > 0$
(ii) there exists some price, $\hat{p} > 0$, at which supply exceeds demand, $S(\hat{p}) > D(\hat{p})$, meaning that $z(\hat{p}) < 0$

then there exists a positive equilibrium price in the market.

Consider the following linear demand and supply curves:

$$D(p) = a - bp, \qquad b > 0$$

$$S(p) = c + ep, \qquad e > 0$$

The discussion above on existence can help us to determine any further conditions on the parameters of these equations (i.e., on a and c) that are needed to guarantee existence of a positive equilibrium price. To generate such an outcome we need to satisfy the conditions listed above in theorem S4.2. Using these assumptions we get the excess-demand function to be

$$z(p) = D(p) - S(p) = (a - c) - (e + b)p$$

The first condition above, that $z(0) > 0$, implies that

$$(a - c) - (e + b)(0) > 0$$

and this shows that we need to restrict the parameters a and c so that $(a - c) > 0$. The second condition requires that there be some price, $\hat{p} > 0$, such that $z(\hat{p}) < 0$. This condition is met by all values of p such that

$$(a - c) - (e + b)p < 0 \Rightarrow p > \frac{a - c}{e + b}$$

With $e > 0$, $b > 0$, and $(a - c) > 0$, we see that this condition is met without any further restrictions on the parameters a and c. Thus a positive equilibrium price is guaranteed to exist for this linear example provided $a > c$.

It is admittedly quite easy to solve explicitly for the equilibrium in an example with linear demand and supply curves. In the example above, $D(p^e) = S(p^e)$, or $z(p^e) = 0$, leads immediately to the result that $p^e = (a - c)/(e + b)$. If $e > 0$ and $b > 0$, it is obvious that $p^e > 0$ only if $a > c$. However, we often work with far less specific models in economics and it is useful to know the general conditions under which a positive equilibrium price will exist. In particular, we see the importance of assuming that the demand and supply functions are continuous. These continuity conditions are implied by specific assumptions about the technology faced by firms and the preferences of consumers. The intermediate-value theorem, if extended to include the case of functions of more than one variable, is also helpful in characterizing the conditions under which a *general equilibrium* or *multi-market equilibrium* will exist. In this case, a simple graph is not so helpful in providing intuition.

Example S4.1 **Supply and Demand Equilibrium**

Consider the following market demand and supply functions:

$$D(p) = 100 - 2p$$

$$S(p) = 3p$$

Graph $D(p)$ and $S(p)$ on one diagram and $z(p) = D(p) - S(p)$ on another. Find the equilibrium price and quantity for this market and illustrate on both graphs. Show that these demand and supply functions satisfy the conditions for the existence of a positive equilibrium price, as specified in theorem S4.2.

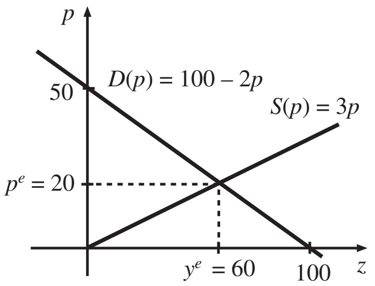

Figure S4.16 Demand and supply functions for example 4.8

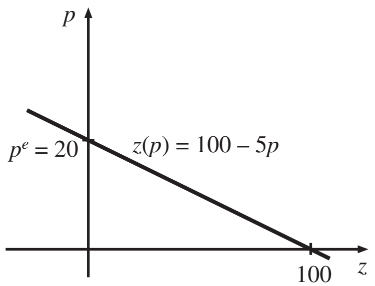

Figure S4.17 Excess-demand function for example 4.8

Solution

Now

$$z(p) = D(p) - S(p) = 100 - 2p - 3p = 100 - 5p$$

The graphs of $D(p)$ and $S(p)$ are shown in figure S4.16 and $z(p)$ is graphed in figure S4.17.

Equilibrium price $p = p^e$ satisfies $D(p^e) = S(p^e)$, so we find it by solving the equation

$$\begin{aligned} 100 - 2p^e &= 3p^e \\ 100 &= 5p^e \\ p^e &= 20 \end{aligned}$$

Substitute $p^e = 20$ back into either the demand or supply function to get equilibrium quantity

$$y^e = D(p^e) = 100 - 2(20) = 60$$

$$y^e = S(p^e) = 3(20) = 60$$

To see that this demand-and-supply system satisfies the requirements of theorem S4.2 note that

(a) at $p = 0$, $D(0) = 100 > S(0) = 0 \Rightarrow z(0) > 0$
(b) at $\hat{p} = 50$, $D(50) = 0 < S(50) = 150 \Rightarrow z(50) < 0$

Since both $D(p)$ and $S(p)$ are linear functions, they are continuous. Therefore all the conditions required for a positive equilibrium price are satisfied. ■

One further matter needs attention. Although for the linear example above, the restrictions imposed on the parameters a, b, c, and e imply that a positive equilibrium price will exist, these conditions do not guarantee that the equilibrium quantity will be nonnegative. Consider the following specific example:

$$D(p) = 10 - 2p$$

$$S(p) = -30 + 3p$$

The solution to this system gives the equilibrium price to be $p^e = 8$ and equilibrium quantity to be $y^e = -6$. If the equilibrium quantity is negative, this means that there is no price that is sufficiently high to induce producers to produce any output, yet at the same time be sufficiently low to induce consumers to purchase any of the commodity. Thus we would say that such a market would not be active; that is, no sales would take place. This is illustrated by figure S4.20. The *choke price,* which is the maximum price at which consumers would be willing to pay for *any* quantity of the output, is 5, while the minimum price that producers must receive in order to be induced to supply any output is 10 and hence this market would not be active. We have also ignored the possibility that there may be more than one equilibrium price that may arise with nonlinear functions.

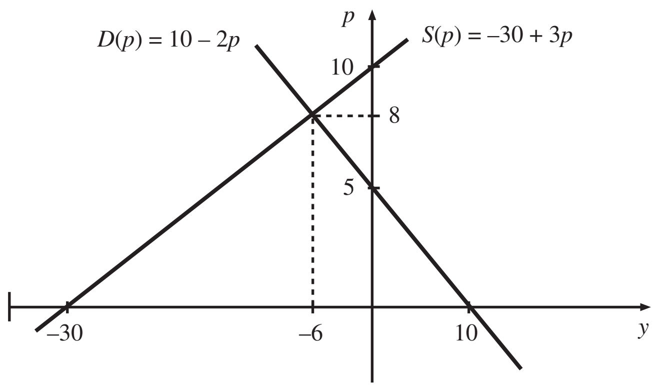

Figure S4.20 A market that would not be active

PRACTICE EXERCISES

S4.1. Let $y = kx$, $k > 0$, be a production function relating input x to output y. Let $\bar{c}$ represent the unit cost of input x, and assume that total cost equals fixed costs, C_0, plus the cost of input x. Let $\bar{p}$ be the unit price of y. Find the revenue function, the cost function, and the profit function for the firm. Are these functions continuous? Discuss. (Use theorem 4.1 to answer this question.)

S4.2. Consider the following example of Hotelling's location model. Each of two firms sells a homogeneous product and charges a price of \$10 while facing constant unit cost of \$6. There are $N = 100$ consumers who are uniformly (evenly) distributed along a street one mile in length, represented by the unit interval [0, 1]. The two firms, A and B, will choose locations, L_A and L_B respectively, on the line [0, 1] in such a way as to maximize market share, and hence profit.

(a) Assuming that firm B establishes its location first at $L_B = 0.3$, find and graph firm A's market-share function $M^A(L_A)$ and profit function $\pi^{A}(L_A)$. Discuss why each function is discontinuous at the point $L_A = 0.3$.

(b) For any general choice of location $L_B < 0.5$ by firm B, do the same exercise as part (a).

(c) For any general choice of location $L_B > 0.5$ by firm B, do the same exercise as part (a).

(d) For choice of location $L_B = 0.5$ by firm B, do the same exercise as part (a).

(e) Since firm B can deduce where firm A would locate, conditional on firm B's own choice of location L_B, where would firm B locate? What is the equilibrium outcome of this model? Discuss.

S4.3. Consider the following market demand and supply functions:

$$D(p) = 50 - 2p$$

$$S(p) = -10 + p$$

Graph $D(p)$ and $S(p)$ on one diagram and $z(p) = D(p) - S(p)$ on another. Find the equilibrium price and quantity for this market and illustrate on both graphs. Show that these demand and supply functions satisfy the requirements for existence of a positive equilibrium price, as specified in theorem S4.2.

S4.4. Consider the following market demand and supply functions:

$$D(p) = a - p, \qquad a > 0$$

$$S(p) = -5 + p$$

(a) What further restriction must we impose, if any, on the parameter a to ensure that a positive equilibrium price exists? (See theorem S4.2.)

(b) What further restriction, if any, must we impose on the parameter a to ensure that an equilibrium with a positive price *and* a positive quantity exists?

S4.5. Consider the following market demand and supply functions:

$$D(p) = 50 - 8p, \qquad S(p) = -100 + 2p$$

Graph $D(p)$ and $S(p)$ on one diagram and $z(p) = D(p) - S(p)$ on another. Find the equilibrium price and quantity for this market and illustrate on both graphs. Show that these demand and supply functions satisfy the conditions for the existence of a positive equilibrium price, as specified in theorem S4.2. Explain what one would actually observe happening in this market.

S4.6. Consider the following market demand and supply functions:

$$D(p) = 20 - p, \quad S(p) = 30 + 4p$$

Graph $D(p)$ and $S(p)$ on one diagram and $z(p) = D(p) - S(p)$ on another. Find the equilibrium price and quantity for this market and illustrate on both graphs. Which of the conditions of theorem S4.2 that guarantee a positive equilibrium price is absent in this example? What would one observe in this market?

S4.7. Consider the following market demand and supply functions:

$$D(p) = 20 + 2p, \quad S(p) = -10 + p$$

Repeat the exercise in question S4.6. Explain what is unusual in this example.

S4.8. Consider the following market demand and supply functions:

$$D(p) = \begin{cases} 50 - 2p, & \text{if } p \geq 10 \\ 70 - 2p, & \text{if } p < 10, \end{cases} \quad S(p) = 10 + 3p$$

Repeat the exercise in question S4.6.

S4.9. Consider the following market demand and supply functions:

$$D(p) = 100 - 2p$$

$$S(p) = -20 + p$$

Graph $D(p)$ and $S(p)$ on one diagram and $z(p) = D(p) - S(p)$ on another. Find the equilibrium price and quantity for this market and illustrate on both graphs. Show that these demand functions satisfy the requirements for existence of a positive equilibrium price, as specified in theorem S4.2.

S4.10. Consider the following market demand and supply functions:

$$D(p) = \begin{cases} 100 - 2p, & p > 45 \\ 120 - 2p, & p \leq 45 \end{cases}$$

$$S(p) = -20 + p$$

Notice that as price falls to \$45, a jump in demand occurs. This may be due to a new group of consumers deciding to enter the market once the price falls to this level.

Graph $D(p)$ and $S(p)$ on one diagram and $z(p) = D(p) - S(p)$ on another. Which conditions of theorem S4.2, that guarantee existence of a (positive) equilibrium price are absent? Discuss.

S4.11. Consider the following example of Hotelling's location model. Each of two firms sells a homogeneous product and charges a price of \$25 while facing constant unit cost of \$15. There are $N = 1{,}000$ consumers who are uniformly (evenly) distributed along a street one mile in length, represented by the unit interval [0, 1]. The two firms, A and B, will choose locations, L_{A} and L_{B} respectively, on the line [0, 1] in such a way as to maximize market share, and hence profit.

(a) Assuming that firm B establishes its location first at $L_{\mathrm{B}} = 0.8$, find and graph firm A's market-share function $M^{\mathrm{A}}(L_{\mathrm{A}})$ and profit function $\pi^{\mathrm{A}}(L_{\mathrm{A}})$. Discuss why each function is discontinuous at point $L_{\mathrm{A}} = 0.8$.

(b) For any general choice of location $L_{\mathrm{B}} < 0.5$ by firm B, do the same exercise as part (a).

(c) For any general choice of location $L_{\mathrm{B}} > 0.5$ by firm B, do the same exercise as part (a).

(d) For choice of location $L_{\mathrm{B}} = 0.5$ by firm B, do the same exercise as part (a).

(e) Since firm B can deduce where firm A would locate, conditional on firm B's own choice of location L_{B}, where would firm B locate? What is the equilibrium outcome of this model? Discuss.

Solutions

S4.1. If $y = kx$ is the production function relating input x to output y, then the inverse of the production function, $x = y/k$, indicates the amount of input x needed in order to produce y units of output. Therefore the variable cost of producing y units of output is

$$VC(y) = \frac{\bar{c}}{k}y$$

Adding in fixed cost, c_0, we get cost function

$$C(y) = c_0 + \frac{\bar{c}}{k}y$$

The revenue function is

$$R(y) = \bar{p}y$$

and the profit function is

$$\pi(y) = R(y) - C(y) = \bar{p}y - c_0 - \frac{\bar{c}}{k}y$$

All of these functions are continuous. One way to see this is to show that the function $y = kx$ is continuous and then note that obtaining the cost function, $C(y)$, involves operations relating to points (vi), (i), and (ii) of theorem 4.1. If we now were to show that $R(y) = \bar{p}y$ is continuous, then obtaining the profit function, $\pi(y)$, involves subtracting one continuous function from another, $(R(y) - C(y))$, which is operation (iii) of theorem 4.1.

S4.2. **(a)** Firm B locates at $L_B = 0.3$. If firm A locates at a point to the left of 0.3 (i.e., $L_A < 0.3$), it will attract all consumers to the left of its choice. This fraction is equal to L_A, plus half of the consumers between L_A and L_B — namely those consumers closer to L_A than L_B — which represents a further fraction of $0.5(L_B - L_A)$. Market share then is

$$M^A(L_A) = L_A + 0.5(0.3 - L_A), \quad L_A < 0.3$$

If firm A chooses to locate at the same point as firm B ($L_A = L_B = 0.3$), then consumers are indifferent between purchasing from A or B and so they share the market equally. Thus

$$M^A(L_A) = 0.5, \quad L_A = 0.3$$

If firm A locates just to the right of firm B, then virtually all consumers to the right of the point 0.3 will go to firm A. This represents the fraction $(1-0.3)$ of the consumers. Further to the right of $L_B = 0.3$, firm A will attract all consumers to the right of its location, representing $1-L_A$ market share, and also 50% of consumers between firm B's location (0.3) and its own location (L_A), thus representing a further fraction $0.5(L_A - 0.3)$ of the market. Thus

$$M^A(L_A) = (1 - L_A) + 0.5(L_A - 0.3), \quad L_A > 0.3$$

Putting these expressions together gives us

$$M^A(L_A) = \begin{cases} L_A + 0.5(0.3 - L_A), & L_A < 0.3 \\ 0.5, & L_A = 0.3 \\ (1 - L_A) + 0.5(L_A - 0.3), & L_A > 0.3 \end{cases}$$

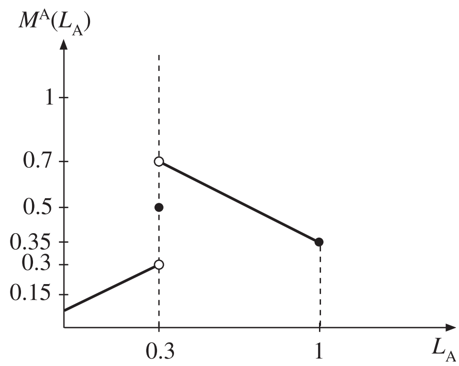

Figure S4.21

There are 100 consumers in the market and the profit made per consumer is \$4 (i.e., \$10 − \$6), and so the profit function relating to market share is

$$\pi^A(L_A) = 400M^A(L_A)$$

By substituting the function for market share into this equation, we get

$$\pi^A(L_A) = \begin{cases} 400[L_A + 0.5(0.3 - L_A)], & L_A < 0.3 \\ 200, & L_A = 0.3 \\ 400[(1 - L_A) + 0.5(L_A - 0.3)], & L_A > 0.3 \end{cases}$$

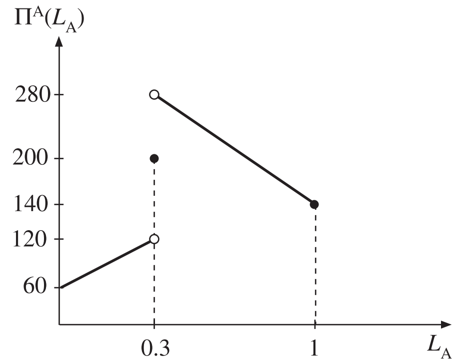

Figure S4.22

Firm A's market share and profit change in a discontinuous fashion at the point $L_A = L_B = 0.3$. Locating just to the left of 0.3 means firm A obtains virtually all the consumers to the left of 0.3, and so

$$\lim_{L_A \to 0.3^-} M^A(L_A) = 0.3$$

is the left-hand limit of $M^A(L_A)$ at the point $L_A = 0.3$. However, if firm A locates precisely at the point $L_A = 0.3$, it shares the market equally with firm B, and so

$$M^A(0.3) = 0.5$$

If firm A locates just to the right of firm B (i.e., to the right of 0.3), it gets all the consumers to the right of 0.3 and so

$$\lim_{L_A \to 0.3^+} M^A(L_A) = 0.7$$

Thus at $L_A = 0.3$ the right-hand limit for $M^A(L_A)$ is not equal to the left-hand limit, and neither is equal to the value of $M^A(L_A)$ at $L_A = 0.3$. Thus, by definition 4.3, we see that this function is not continuous at this point.

(b) The same arguments made in part (a) apply to any choice $\bar{L}_B < 0.5$ made by firm B. Firm A's market share function is

$$M^A(L_A) = \begin{cases} L_A + 0.5(\bar{L}_B - L_A), & L_A < \bar{L}_B \\ 0.5, & L_A = \bar{L}_B \\ (1 - L_A) + 0.5(L_A - \bar{L}_B), & L_A > \bar{L}_B \end{cases}$$

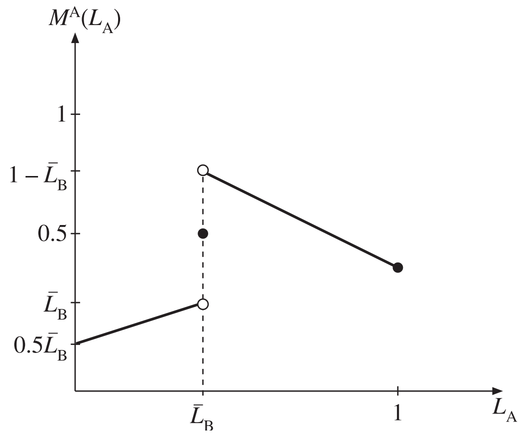

Figure S4.23

and the profit share function is

$$\pi^A(L_A) = \begin{cases} 400[L_A + 0.5(\bar{L}_B - L_A)], & L_A < \bar{L}_B \\ 200, & L_A = \bar{L}_B \\ 400[(1 - L_A) + 0.5(L_A - \bar{L}_B)], & L_A > \bar{L}_B \end{cases}$$

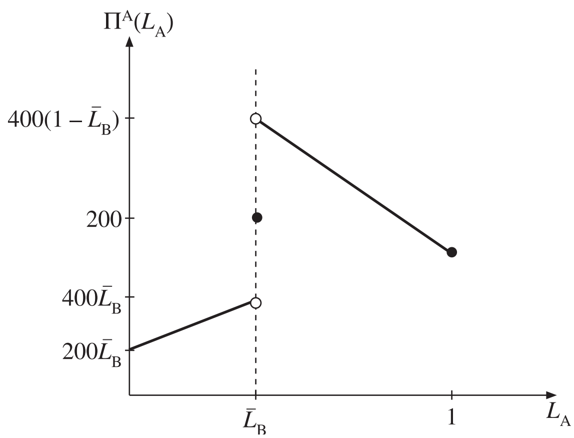

Figure S4.24

(c) If firm B locates to the right of midpoint ($\bar{L}_B > 0.5$), then a similar method is used to determine A's market share and profit as was used in part (b). The one important difference is that firm A's market share and profit are higher when it locates just to the left of firm B's location than if it locates just to the right. Thus the same functions as in part (b) describe the market share and profit for firm A, but the appropriate graphs are in figures S4.25 and S4.26.

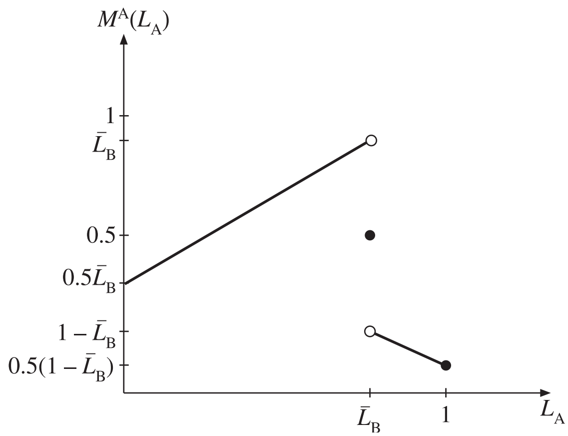

Figure S4.25

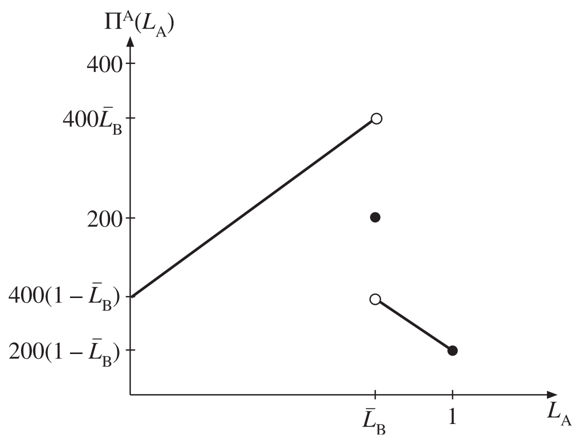

Figure S4.26

(d) Once again, the functions in part (b) describe the market share and profit for firm A as a function of its location, L_A. However, since $\bar{L}_B = 0.5$, firm A's market share approaches the value 0.5 as L_A approaches the value 0.5 either from the left ($L_A < 0.5$) or from the right ($L_A > 0.5$). Moreover market share for A is 0.5 when it locates at the point $L_A = \bar{L}_B = 0.5$. Therefore the market share function is continuous at the point $L_A = \bar{L}_B$, as is the profit function. The appropriate graphs are given in figures S4.27 and S4.28.

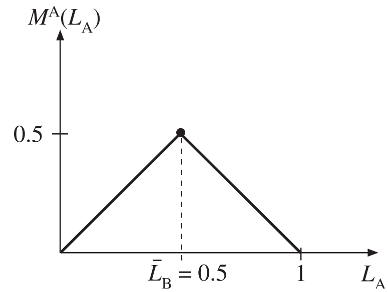

Figure S4.27

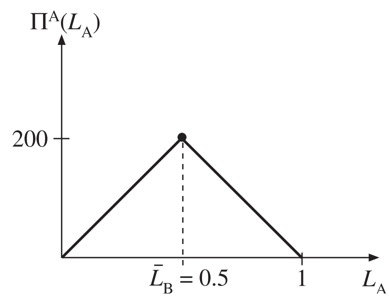

Figure S4.28

(e) Firm B knows that if it locates either to the left or right of the midpoint, 0.5, then firm A will choose its location so that firm A obtains greater than 50% of the market share. This leaves firm B with less than 50% of the market share. Therefore the best that firm B can do is to locate at the midpoint, recognizing that firm A will respond by also locating at the midpoint, with the firms sharing the market equally. The reverse argument also applies if we consider where firm A would locate based on the reaction of firm B, and so the equilibrium outcome is that both firms locate at the midpoint.

S4.3. To find the equilibrium price, set $D(p) = S(p)$:

$$\begin{aligned} 50 - 2p^e &= -10 + p^e \\ -3p^e &= -60 \\ p^e &= 20 \end{aligned}$$

To find the equilibrium quantity, substitute $p^e = 20$ into either the demand or supply function (since they are equal at the equilibrium price).

$$D(p^e) = 50 - 2p^e \Longrightarrow y^e = 10$$

Also

$$S(p^e) = -10 + p^e \Longrightarrow y^e = 10$$

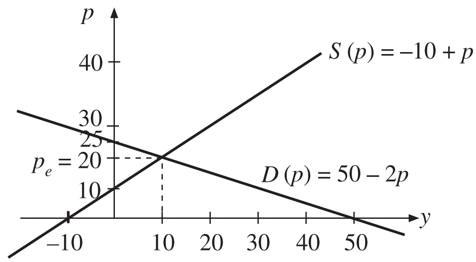

Figure S4.29

The excess demand function is

$$\begin{aligned} z(p) &= D(p) - S(p) = (50 - 2p) - (-10 + p) \\ &= 60 - 3p \end{aligned}$$

Note that $z(p^e) = 0 \Longrightarrow p^e = 20$ is another method of determining the equilibrium price.

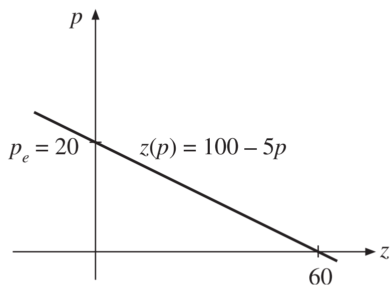

Figure S4.30

The demand and supply functions satisfy theorem S4.2, since:
(i) $D(0) = 50$, $S(0) = -10$, and so $D(0) > S(0)$.
(ii) Picking price $\bar{p} = 25$ (or any price exceeding 20 in this case), we see that there is indeed a price $p = \bar{p}$ such that $S(\bar{p}) > D(\bar{p})$ (i.e., note that $S(25) = 15$, $D(25) = 0$).

S4.4. **(a)** Theorem S4.2 requires that
(i) $D(0) > S(0)$
(ii) There is some price $\bar{p}$ such that $S(\bar{p}) > D(\bar{p})$. Since $D(0) = a$ and $S(0) = -5$, the condition $a > 0$ implies that $D(0) > S(0)$ and $S(\bar{p}) > D(\bar{p})$ for any $\bar{p}$ such that $-5 + \bar{p} > a - \bar{p} \Rightarrow 2\bar{p} > a + 5 \Rightarrow \bar{p} > (a+5)/2$, and so such a $\bar{p}$ exists for any $a > 0$.

(b)

$$\begin{aligned} z(p) &= D(p) - S(p) = (a - p) - (-5 + p) \\ &= (a + 5) - 2p \end{aligned}$$

In equilibrium, $z(p^e) = 0$, and so

$$(a + 5) - 2p^e = 0 \Rightarrow p^e = \frac{a + 5}{2}$$

Substituting p^e into the demand (or supply) function gives the equilibrium quantity:

$$D(p^e) = a - p^e = a - \frac{a + 5}{2} = \frac{a}{2} - 2.5$$

That is,

$$y^e = \frac{a}{2} - 2.5$$

So $y^e > 0$ only if $a/2 - 2.5 > 0 \Longrightarrow a > 5$. Consequently the further restriction on a of $a > 5$ must be imposed for there to be a positive equilibrium quantity.

S4.5. $D(p) = 50 - 8p$ is the demand function. $S(p) = -100 + 2p$ is the supply function.

$$z(p) = D(p) - S(p) = (50 - 8p) - (-100 + 2p)$$

and so $z(p) = 150 - 10p$ is the excess demand function.

The equilibrium price, p^e, satisfies $z(p^e) = 0$, and so $p^e = 15$. To find the equilibrium quantity, substitute p^e into either $D(p)$ or $S(p)$. We get $y^e = -70$.

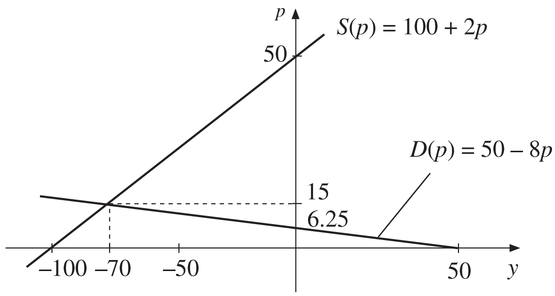

Figure S4.31

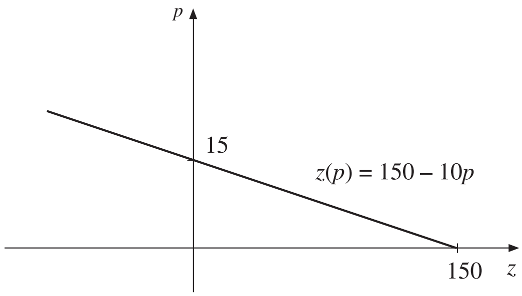

Figure S4.32

Since $D(0) = 50$ and $S(0) = -100$ where $D(0) > S(0)$, condition (a) of theorem 4.3 is satisfied. For any $\bar{p} > 15$ we have $D(\bar{p}) < S(\bar{p})$ or $z(\bar{p}) < 0$, and so condition (b) of theorem S4.2 is also satisfied. What is peculiar about this example is that using the standard computational approach delivers a negative value for the equilibrium quantity, which makes no economic sense. Upon considering the vertical (p) intercepts for both the supply and demand functions, we see that firms require a minimum price of \$50 in order to be induced to produce any of this product while consumers will not purchase any output if price exceeds \$6.25. Therefore there is no price that will induce any (positive) market transactions. As a result this market will be inactive.

S4.6. $D(p) = 20 - p$ is the demand function. $S(p) = 30 + 4p$ is the supply function.

$$z(p) = D(p) - S(p) = (20 - p) - (30 + 4p)$$

and so $z(p) = -10 - 5p$ is the excess demand function.

The equilibrium price, p^e, satisfies $z(p^e) = 0$ and so $p^e = -2$. To find the equilibrium quantity, substitute p^e into either $D(p)$ or $S(p)$. We get $y^e = 22$.

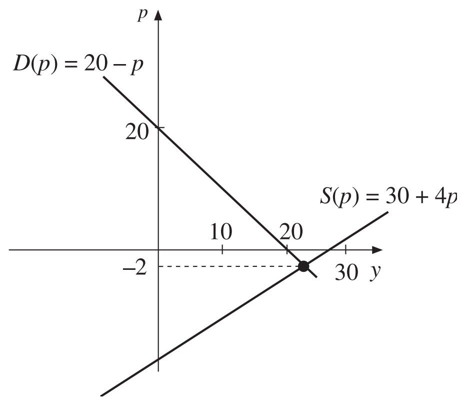

Figure S4.33

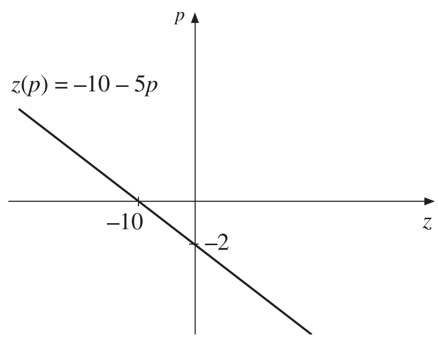

Figure S4.34

Since $D(0) = 20$ and $S(0) = 30$, the first condition of theorem S4.2, that $D(0) > S(0)$, is not satisfied. This means that if the price were zero, consumers would only wish to consume 20 units of the good while 30 units would be available. Therefore this is an example of a free good.

S4.7. $D(p) = 20 + 2p$, $S(p) = -10 + p$

$$z(p) = D(p) - S(p) = (20 + 2p) - (-10 + p)$$

and so $z(p) = 30 + p$ is the excess demand function.

The equilibrium price, p^e, satisfies $z(p^e) = 0$ and so $p^e = -30$. To find the equilibrium quantity, substitute p^e into either $D(p)$ or $S(p)$. We get $y^e = -40$.

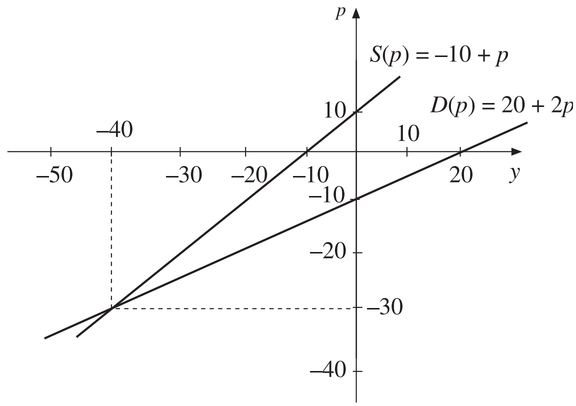

Figure S4.35

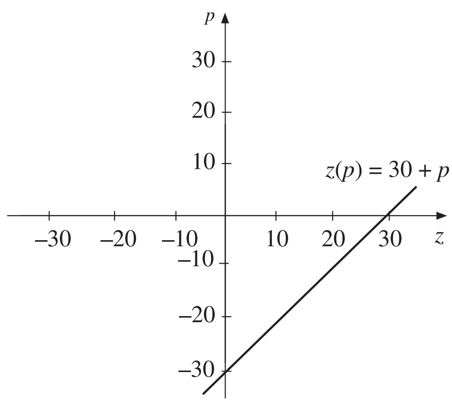

Figure S4.36

Since $D(0) = 20$ and $S(0) = -10$, the first condition of theorem S4.2, that $D(0) > S(0)$, is satisfied. However, since $S(p) > D(p)$ implies that $-10 + p > 20 + 2p$ or $p < 30$, there is no positive price $\bar{p}$

such that $S(\bar{p}) > D(\bar{p})$. So the second condition of theorem S4.2 is not satisfied.

In this example, for any price greater than \$10, a positive amount of output will be supplied. However, at every price greater than \$10, demand exceeds supply, and so firms could increase price and still sell their entire output. One would expect to see a price increasing without bound. Of course, a commodity with demand increasing in price without bound is not a realistic possibility, since consumers would exhaust their incomes if such were the case.

S4.8.

$$D(p) = \begin{cases} 50 - 2p & \text{if } p \geq 10 \\ 70 - 2p & \text{if } p < 10 \end{cases}$$

$$S(p) = 10 + 3p$$

$z(p) = D(p) - S(p)$, and so the excess demand function is

$$z(p) = \begin{cases} 50 - 2p - (10 + 3p) = 40 - 5p & \text{if } p \geq 10 \\ 70 - 2p - (10 + 3p) = 60 - 5p & \text{if } p < 10 \end{cases}$$

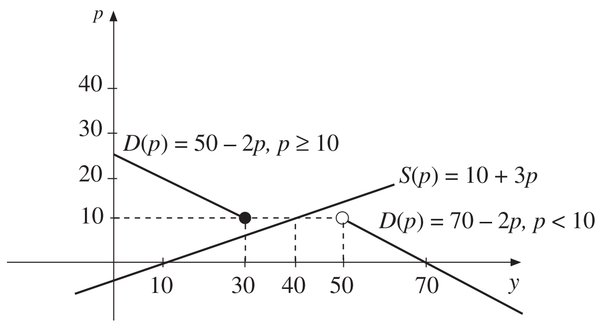

Figure S4.37

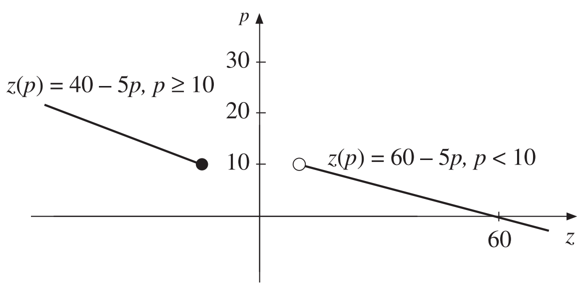

Figure S4.38

We can see from the graphs that the demand function is discontinuous at $p = 10$, as is the excess demand function. Therefore theorem S4.2 does not apply. As it turns out, there is no price that equates demand and supply. If $p \geq 10$, then we have $D(p) < S(p)$. However, if $p < 10$, then we have $D(p) > S(p)$. Although one might expect to see a price of \$10 in this market, at least some firms would discover at this price that some of their output was not being sold, and so inventories would build up. If price was reduced even a small amount, demand would then actually outstrip supply. Other than this observation, the equilibrim model does not, in this case, indicate clearly how such a market would behave.

S4.9. Letting y represent quantity for either the demand or supply function and z the excess demand, we have $D(p) = 100-2p$, $S(p) = -20+p$, $z(p) = (100-2p)-(-20+p) \Rightarrow z(p) = 120-3p$. Equilibrium price, p^e, is determined by $D(p^e) = S(p^e)$ or $z(p^e) = 0$, and so $p^e = 40$. Substitution of $p^e = 40$ into either the demand or supply function gives us $y^e = 20$.

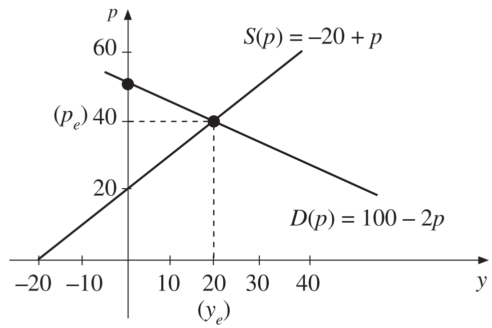

Figure S4.39

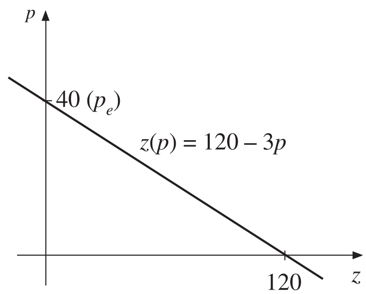

Figure S4.40

The demand and supply functions are continuous, and as a result so is the excess demand function. Since $D(0) = 100 > S(0) = -20$, condition (i) of theorem S4.2 is satisfied. Also, since $D(p) < S(p) \iff 100 - 2p < -20 + p \iff p > 40$, it follows that condition (ii) of theorem S4.2 is satisfied (i.e., $D(\bar{p}) < S(\bar{p})$ for any value $\bar{p} > 40$).

S4.10. Letting y represent quantity for either the demand or supply function and z the excess demand, we have

$$D(p) = \begin{cases} 100 - 2p, & p > 45 \\ 120 - 2p, & p \leq 45 \end{cases}$$

$$S(p) = -20 + p$$

$$z(p) = \begin{cases} 120 - 3p, & p > 45 \\ 140 - 3p, & p \leq 45 \end{cases}$$

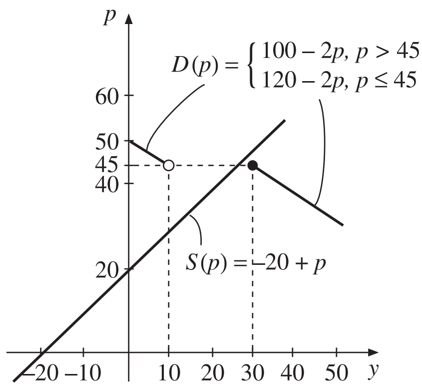

Figure S4.41

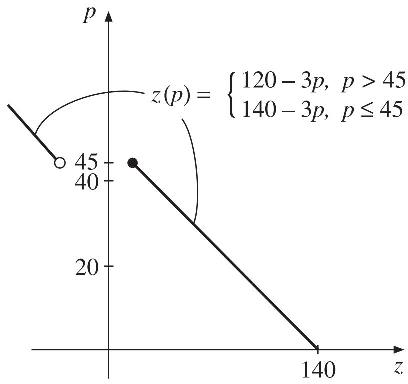

Figure S4.42

Since the demand function is not continuous, and hence neither is the excess demand function, theorem S4.2 does not apply, and so there may not be an equilibrium. This is seen to be the case for this exercise, as is illustrated by the graphs in figures S4.41 and S4.42.

S4.11. This exercise is similar to exercise S4.2. The numbers are different, and notice that when firm B chooses a location to the right of the midpoint ($L_B = 0.8$), firm A's largest market share is obtained when it locates just to the *left* of 0.8, as indicated in figure S4.43. Due to the similarity of these questions, we indicate below only how they differ.

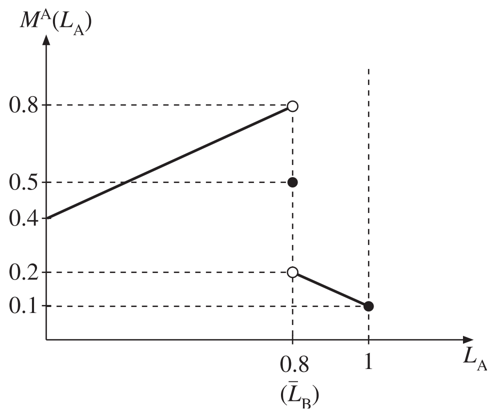

Figure S4.43

(a)

$$M^{\mathrm{A}}(L_{\mathrm{A}}) = \begin{cases} L_{\mathrm{A}} + 0.5(0.8 - L_{\mathrm{A}}), & L_{\mathrm{A}} < 0.8 \\ 0.5, & L_{\mathrm{A}} = 0.8 \\ (1 - L_{\mathrm{A}}) + 0.5(L_{\mathrm{A}} - 0.8), & L_{\mathrm{A}} > 0.8 \end{cases}$$

Since $\pi^{\mathrm{A}}(L_{\mathrm{A}}) = 10,000M^{\mathrm{A}}(L_{\mathrm{A}})$,

$$\pi^{\mathrm{A}}(L_{\mathrm{A}}) = \begin{cases} 10{,}000[L_{\mathrm{A}} + 0.5(0.8 - L_{\mathrm{A}})], & L_{\mathrm{A}} < 0.8 \\ 5{,}000, & L_{\mathrm{A}} = 0.8 \\ 10{,}000[(1 - L_{\mathrm{A}}) + 0.5(L_{\mathrm{A}} - 0.8)], & L_{\mathrm{A}} > 0.8 \end{cases}$$

(b) Use the same functions as in (a), but replace 0.8 by $\bar{L}_{\mathrm{B}}$ everywhere. Notice that when $\bar{L}_{\mathrm{B}} < 0.5$, firm A's market share and profit are highest if it locates just to the right of $\bar{L}_{\mathrm{B}}$

(c) The functions are the same functions as in (b) except that firm A's market share and profit are highest if it locates just to the left of $\bar{L}_{\mathrm{B}}$.

(d) The functions are the same functions as in parts (b) and (c) except that firm A's market share and profit are highest if it locates just at $L_{\mathrm{A}} = \bar{L}_{\mathrm{B}} = 0.5$, and the market share and profit functions are continuous at this point.

(e) Firm B would locate at $L_{\mathrm{B}} = 0.5$, and this would be the equilibrium outcome of the model.

Chapter S5 The Derivative and Differential for Functions of One Variable

Contents

S5.1 Marginal Revenue Product of Labor for a Competitive Firm and for a Monopoly Firm

Let $q = q(L)$ be a firm's production function where q is output and L is a single input of labor. The **marginal (physical) product of labor** is $\mathrm{MP}(L) = dq/dL$ and measures the extra units of output resulting from one more unit of input, L. The **marginal value product of labor,** $\mathrm{MVP}(L)$, is the market value of the output created by employing one more unit of input, which is just the amount of extra output multiplied by the price. Thus we can write

$$\mathrm{MVP}(L) = p\mathrm{MP}(L)$$

where p is the price of the product.

The added revenue to the firm which results from using an additional unit of labor is called the **marginal revenue product of labor,** $\mathrm{MRP}(L)$. If $R(q)$ is the firm's revenue function, then using the chain rule we can write

$$\mathrm{MRP}(L) = \frac{dR}{dL} = \frac{dR}{dq}\frac{dq}{dL} = \mathrm{MR}(q)\mathrm{MP}(L)$$

We saw in example 5.4 that a competitive firm's marginal revenue function is simply equal to market price, $\mathrm{MR}(q) = \bar{p}$. It follows then that $\mathrm{MVP}(L) = \mathrm{MRP}(L)$ for

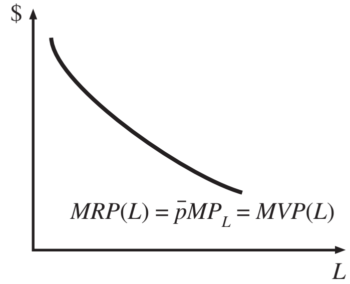

Figure S5.1 Marginal revenue product and marginal value product coincide for a competitive firm

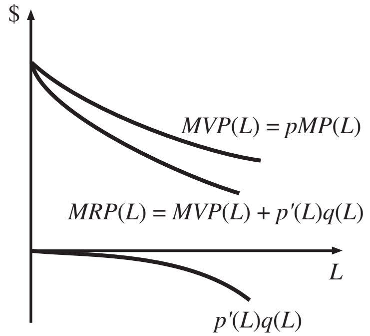

Figure S5.2 Relationship between marginal revenue product and marginal value product for a monopoly firm

a competitive firm. Assuming that $\mathrm{MP}(L)$ is decreasing in L, then this relationship is illustrated in figure S5.1.

For a monopoly firm, however, price depends on the level of output produced and hence depends on the level of the input used. If we write the inverse demand function as $p = p(q)$ or $p = p(q(L))$ and assume $p'(q) < 0$, then we can write the monopolist's revenue function in terms of its input L as

$$R(q(L)) = p(q(L))q(L)$$

where the term $p(q(L))$ is to be read as "p is a function of q which is a function of L." Thus, according to the chain rule, we get

$$p'(L) \text{ or } \frac{dp}{dL} = \frac{dp}{dq}\frac{dq}{dL}$$

Using these results we get the following expression for the marginal revenue product of labor for a monopoly firm:

$$\mathrm{MRP}(L) = \frac{dR}{dL} = \frac{dp}{dq}\frac{dq}{dL}q + \frac{dq}{dL}p = p'(L)q + \mathrm{MP}(L)p$$

Given that $dp/dq < 0$ and $dq/dL > 0$, it follows that $p'(L) = (dp/dq)(dq/dL) < 0$ and so $\mathrm{MRP}(L) < \mathrm{MP}(L)p = \mathrm{MVP}(L)$. Therefore the monopolist's marginal revenue product of labor curve lies below the marginal value product of labor curve. Intuitively speaking, the monopolist values the use of more labor less than the market does (i.e., less than consumers do). This is illustrated in figure S5.2 and by the following example.

Example S5.1

Marginal Revenue Product of Labor

Show that the marginal revenue product of labor, $\mathrm{MRP}(L)$, for a monopolist with inverse demand function $p(q) = 30 - q$ and production function $q(L) = 2L^{1/2}$ is less than the marginal value product of labor $\mathrm{MVP}(L)$, where $\mathrm{MVP}(L) = p\mathrm{MP}(L)$.

Solution

We have

$$\mathrm{MP}(L) = \frac{dq}{dL} = L^{-1/2}$$

and so

$$\text{MVP}(L) = p\text{MP}(L) = pL^{-1/2}$$

On substituting from the inverse demand function, and the production function, we can write p as a function of labor

$$p(L) = 30 - q = 30 - 2L^{1/2}$$

which implies that

$$\text{MVP}(L) = (30 - 2L^{1/2})L^{-1/2} = 30L^{-1/2} - 2$$

The monopolist's revenue, written as a function of labor, is

$$\begin{aligned} R(q(L)) &= p(q(L))q(L) = (30 - q(L))q(L) = (30 - 2L^{1/2})(2L^{1/2}) \\ &= 60L^{1/2} - 4L \end{aligned}$$

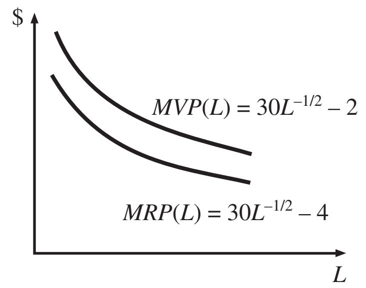

Figure S5.3 Graph for example S5.1

Therefore

$$\text{MRP}(L) = \frac{dR}{dL} = 30L^{-1/2} - 4$$

which is less than

$$\text{MVP}(L) = 30L^{-1/2} - 2$$

See figure S5.3. ■

S5.2 Further Details on the Elasticity Concept

For the example in the text, the arc elasticity between points (y_1, p_1) and (y_2, p_2) is

$$\left[-\frac{\%\Delta y}{\%\Delta p}\right] = -\frac{-28.6}{66.7} \doteq 0.43$$

that is, 43%. This elasticity represents, in absolute value, the percentage change in demand (y) due to a given percentage change in price (see figure S5.4).

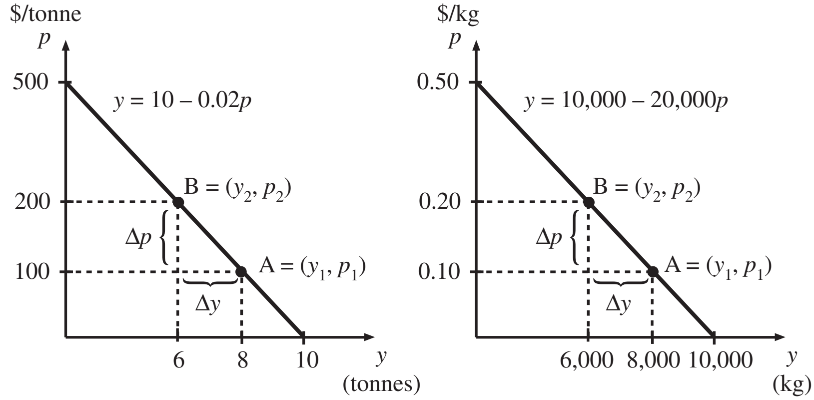

Figure S5.4 Single demand relationship expressed in terms of different units

One of the problems with arc elasticity is that the value depends on the size of the price change, Δp. Take the same initial point as in the example above, $p_1 = \$100$ per tonne and $y_1 = 8$ tonnes, but use $\Delta p = \$50$ to get $p_2 = \$150$ per tonne and $y_2 = 7$ tonnes. In this case the arc elasticity is

$$\left[-\frac{\%\Delta y}{\%\Delta p}\right] = -\frac{\Delta y/(y_1 + y_2)}{\Delta p/(p_1 + p_2)} = -\frac{-1/(7+8)}{50/(100+150)} = \frac{5}{15} \doteq 0.33$$

Since there is no obvious amount by which one should change price to determine the elasticity value, the concept of arc elasticity is troublesome. The source of the difficulty is that the arc elasticity is really an average elasticity between the points (y_1, p_1) and (y_2, p_2). By taking the limit of the arc elasticity formula as $\Delta p \to 0$, we get the two points converging and in so doing we find the elasticity *at the point* (y_1, p_1). The **point elasticity of demand** is developed formally in the text.

Example S5.2 Finding Elasticities of Demand

Return to our example in the text a demand function for steel. Using the first form (units of measurement) of the demand function, $y = 10 - 0.02p$, the point elasticity of demand at $(y_1, p_1) = (8, 100)$ is

$$\epsilon = -\frac{dy}{dp}\frac{p_1}{y_1} = -\left[-0.02\left(\frac{100}{8}\right)\right] = \frac{2}{8} = 0.25$$

Using the other form, $y = 10{,}000 - 20{,}000p$, the point elasticity of demand at the equivalent point, $(y_1, p_1) = (8{,}000, 0.1)$, is

$$\epsilon = -\frac{dy}{dp}\frac{p_1}{y_1} = -\left[-20{,}000\left(\frac{0.1}{8{,}000}\right)\right] = 0.25$$

We see that the choice of units of measurement does not affect the point elasticity value. ■

In figure S5.5 we see that the elasticity of demand falls as one moves from point A to point B. Since at point A the base level of y is *small,* a small percentage reduction in price leads to a large percentage increase in y. Alternatively, as one moves towards point B, the percentage increase in y becomes smaller as the base level of y rises and the base level of p falls. This is seen mathematically from the fact that the slope (dy/dp) is constant along a linear demand curve but the ratio p/y falls as one moves from point A to point B. In fact, upon shifting point A so that it approaches the p intercept we have $y \to 0$ and so $\epsilon \to \infty$, while, upon shifting point B so that it approaches the y intercept, we have $p \to 0$ and so $\epsilon \to 0$. At the midpoint (i.e., the midpoint between $y = a$, $p = 0$ and $y = 0$, $p = a/b$) we have that $y = a/2$, $p = a/2b$, and so

$$\epsilon = b\left(\frac{a/2b}{a/2}\right) = 1$$

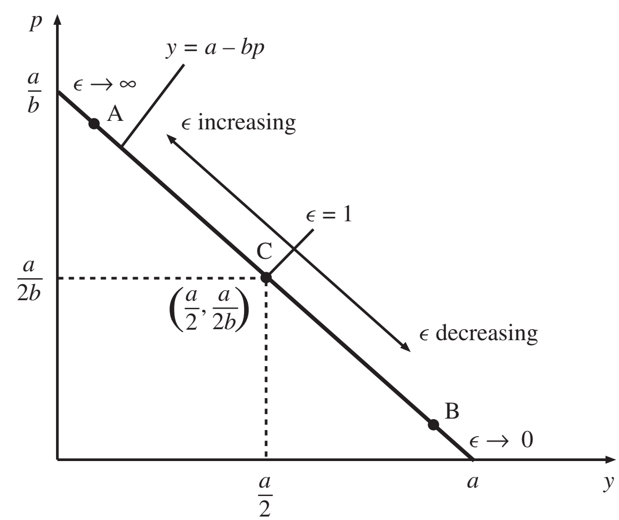

Figure S5.5 Elasticity changes along a linear demand curve

Another functional form which is frequently used to express the relationship between price and quantity demanded is the so-called **constant elasticity demand function:**

$$y = \alpha p^{-\beta}, \quad \alpha, \beta > 0$$

Consider first the particular case with $\beta = 1$ (i.e., $y = \alpha/p$), constant unitary elasticity of demand. For this demand function it is easy to see that total sales revenue is the same at any price ($py = \alpha$). This is illustrated in figure S5.6. Since total sales revenue is constant at any price, then it follows that any percentage reduction (increase) in price is always matched by the same percentage increase (reduction) in sales, which keeps total sales revenue constant. Thus the elasticity of demand is a constant value equal to one at every point on the demand function. This is seen by applying the formula $\epsilon = -(dy/dp)(p/y)$ to the function $y = \alpha p^{-1}$ to get

$$\epsilon = -((-1)\alpha p^{-2})\frac{p}{y}$$

Upon substitution for $y = \alpha p^{-1}$, we get

$$\epsilon = \frac{\alpha p^{-1}}{\alpha p^{-1}} = 1$$

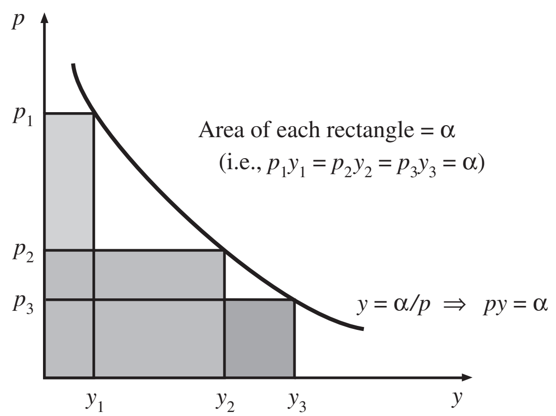

Figure S5.6 Constant unitary elasticity of demand function

For the more general form $y = \alpha p^{-\beta}$, we get the following result:

$$\epsilon = -\frac{dy}{dp}\frac{p}{y} = -((-\beta)\alpha p^{-\beta-1})\frac{p}{y} = \frac{\alpha\beta p^{-\beta}}{\alpha p^{-\beta}} = \beta$$

PRACTICE EXERCISES

S5.1. Suppose that a monopoly firm faces the demand function $q = 10{,}000 - 200p$. Consider an initial price and output combination of $\hat{p} = 10$ and $\hat{q} = 8{,}000$.

(a) Find the extra revenue generated from the extra sales that would result from a change in the price $\Delta p = -1$ (i.e., $\Delta q = 200$).

(b) Find the loss of revenue caused by the units that would have been made at price $\hat{p} = 10$ but are now sold at price $p = 9$.

(c) Illustrate the two values computed in parts (a) and (b) on a diagram such as figure 5.24, and show how these two values make up the change in revenue associated with increased sales of $\Delta q = 200$.

(d) Find the marginal revenue function for this monopolist.

Solution

S5.1. **(a)** If the price is reduced from $p = \$10$ to $p = \$9$ ($\Delta p = -1$), sales increase from $q = 8{,}000$ to $q = 8{,}200$ ($\Delta q = 200$) and the sales revenue from these additional sales is $200 \times \$9 = \$1{,}800$ (see area A in figure S5.7).

(b) The 8,000 units that could have been sold at price $p = \$10$ but are now sold at $p = \$9$ generate \$8,000 less in revenue (see area B in figure S5.7).

(c) See figure S5.7

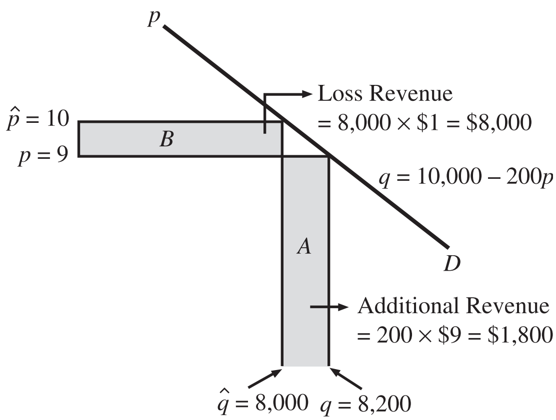

Figure S5.7

(d) The inverse demand function for this monopolist is

$$p = 50 - (1/200)q$$

and so the total revenue function is

$$\mathrm{TR}(q) = pq = [50 - (1/200)q]q = 50q - (1/200)q^2$$

Thus the marginal revenue function is

$$\mathrm{MR}(q) = \mathrm{TR}'(q) = 50 - (1/100)q$$

Chapter S6 Optimization of Functions of One Variable

Contents

S6.1 Monopoly Equilibria I and II

Example S6.1 **Monopoly Equilibrium I**

A monopolist has inverse demand function $p = 50 - 2x$. The total-cost function is $C = 20 + 2x + 0.5x^2$. What are the profit-maximizing price and output?

Solution

Profit is

$$\begin{aligned}\pi(x) &= 50x - 2x^2 - [20 + 2x + 0.5x^2] \\ &= 48x - 2.5x^2 - 20\end{aligned}$$

so that

$$\pi'(x^*) = 48 - 5x^* = 0$$

and so

$$x^* = 9.6 \quad p^* = 50 - 2(9.6) = \$30.80$$

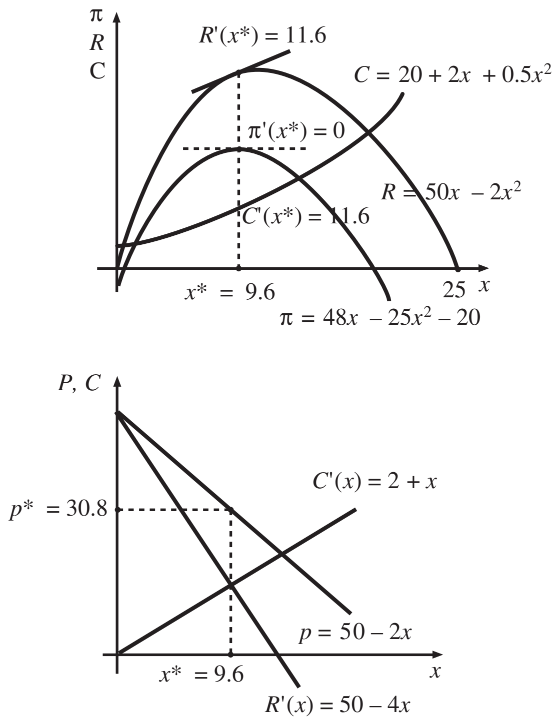

Figure S6.1 Monopoly equilibrium for example S6.2

The level of profit at the maximum is then

$$\pi(x^*) = 48(9.6) - 2.5(9.6)^2 - 20 = \$210.40$$

See figure S6.1. ■

Example S6.2 **Monopoly Equilibrium II**

A monopolist has inverse demand function $p = 150 - 2x$ and total-cost function $C = 0.1x^3 - 3x^2 + 50x + 100$. What are the profit-maximizing price and output?

Solution

A similar sequence of steps to those in exercise 6.1 gives

$$\begin{aligned}\pi(x) &= 150x - 2x^2 - [0.1x^3 - 3x^2 + 50x + 100] \\ &= 100x + x^2 - 0.1x^3 - 100\end{aligned}$$

$$\pi'(x^*) = 100 + 2x - 0.3x^2 = 0$$

Solving this quadratic gives

$$x^* = \frac{-2 \pm \sqrt{4 - 4(-0.3)(100)}}{-0.6}$$

so that

$$x^* = 21.89 \quad \text{or} \quad x^* = -15.23$$

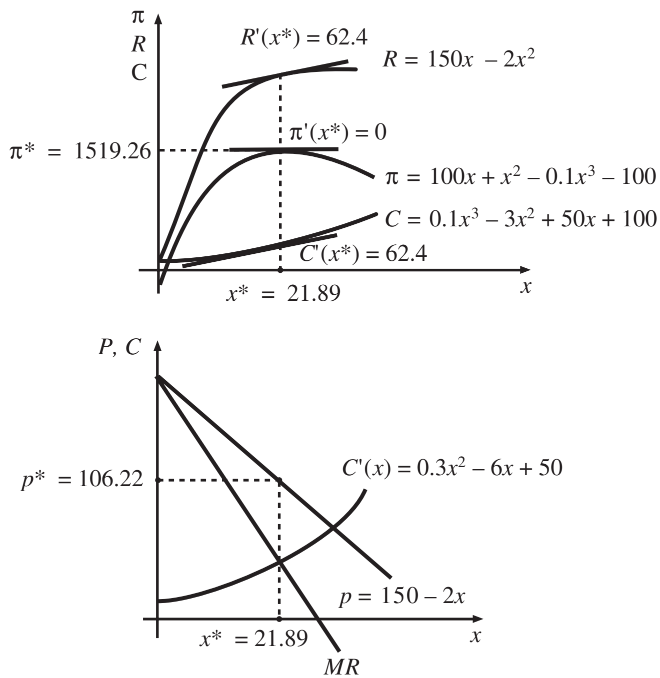

Figure S6.2 Monopoly equilibrium for example S6.3

Since negative outputs are impossible, $x^* = 21.89$, $p^* = \$106.22$, and

$$\pi(x^*) = 100(21.89) + (21.89)^2 - 0.1(21.89)^3 - 100 = \$1{,}519.26$$

See figure S6.2. ■

S6.2 Monopoly with Constant-Elasticity Demand and Constant Costs

We begin with an example of constant *unit* elasticity of demand. Suppose that a monopoly firm faces the demand function $x = 10p^{-1}$, or in inverse form $p = 10x^{-1}$, and has the cost function, $C = 5x$. Note that we have to restrict the domain of the function to $x > 0$. The demand function is called "constant elasticity" because, if we evaluate the elasticity of demand, we obtain

$$\epsilon = -\frac{dx}{dp}\frac{p}{x} = -(-10p^{-2})\left(\frac{p}{10p^{-1}}\right) = 1$$

or even more simply, since $\log x = \log 10 - \log p$

$$\epsilon = -\frac{d \log x}{d \log p} = 1$$

Since the elasticity is independent of the particular point on the demand curve, we say elasticity is *constant*.

To find the profit-maximizing output, we set up

$$R(x) = px = (10x^{-1})x = 10$$

$$C(x) = 5x$$

$$\pi(x) = R(x) - C(x) = 10 - 5x$$

and applying equation (6.3) gives

$$\pi'(x^*) = -5 = 0$$

which again is nonsense. What went wrong this time? Again, a figure will suggest the mathematical answer (see figure S6.3). In figure S6.3(a), the revenue function $R(x)$ is a constant, while costs are increasing, and so profit varies inversely with output—the lower is output the higher is profit.

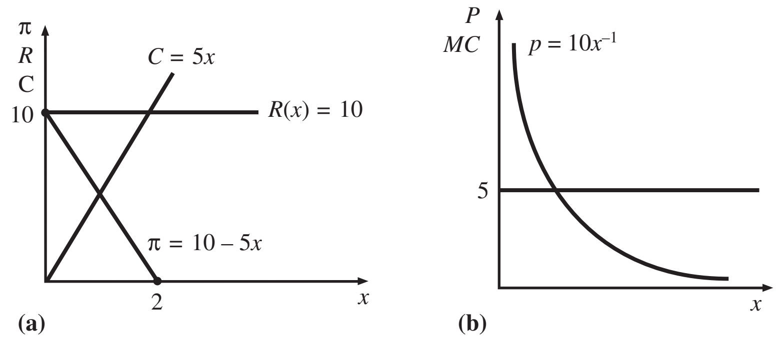

Figure S6.3 Monopoly with unit elastic demand

In figure S6.3(b), we see why this is so: as output is lowered, price rises by enough to keep revenue constant ($p = 10/x \Rightarrow px = 10$), and so, since reducing output leaves revenue unchanged while reducing total costs, it pays to produce as small an output as possible. However, there is a problem here. If the firm produced zero output, it makes zero profit, so this cannot be the maximum. But then, since between zero and any x, however small, there is an infinity of x-values (x is a real number), so there is in fact *no solution to the problem*. If we proposed, say, $x = 0.1$ as a solution, we could immediately show that $x = 0.01$ yielded a larger profit, $x = 0.001$ still more, and so on.

Another way of looking at this is to note that marginal revenue, $R'(x)$, in this case is zero: in figure S6.3(b), the marginal-revenue "curve" coincides with the horizontal axis, since a change in output produces no change in revenue. Since marginal cost is constant at \$5, there can be no output at which marginal revenue equals marginal cost. Again, we see that care must be taken in applying equation (6.3) in a model, because we again have a case in which a maximum solution does not exist. A general discussion of the existence of solutions to optimization problems is presented in chapter 13.

Now consider an example with constant elasticity of demand which is greater than one. Let $x = p^{-2}$ be the demand curve faced by the monopoly. In inverse form we express this demand function as $p = x^{-1/2}$. The elasticity of demand is

$$\epsilon = -\frac{dx}{dp}\frac{p}{x} = -(-2p^{-3})\left(\frac{p}{p^{-2}}\right) = 2$$

In this case the revenue function is not a constant:

$$R(x) = px = x^{-1/2}x = x^{1/2}$$

If the cost function is $C(x) = 2x$, we have profit given by

$$\pi(x) = R(x) - C(x) = x^{1/2} - 2x$$

and so

$$\pi'(x^*) = 0 \Rightarrow \frac{x^{*-1/2}}{2} - 2 = 0$$

so that $x^* = 1/16$ is the level of output that will maximize profit. Unlike the case of unit elasticity of demand, we see in this example that reducing output reduces total revenue and so a positive value of output exists which gives maximum profit.

The third possibility is for the constant elasticity of demand to be less than one. We treat this possibility below.

Example S6.3 Constant Elasticity of Demand Less Than One

Is there a level of output, $x \geq 0$, which maximizes profit for a monopolist facing the demand function $x = p^{-1/2}$ and cost function $C(x) = 2x$? Discuss in terms of the elasticity of demand.

Solution

Since

$$x = p^{-1/2} \Rightarrow x = \frac{1}{p^{1/2}} \Rightarrow p^{1/2}x = 1 \Rightarrow p^{1/2} = x^{-1}$$

we get $p = x^{-2}$ as the inverse demand function. We also have

$$\epsilon = -\frac{dx}{dp}\frac{p}{x} = -\left(-\frac{1}{2}p^{-3/2}\right)\frac{p}{p^{-1/2}} = \frac{1}{2}$$

Thus elasticity is a constant, less than one. Moreover

$$R(x) = px = x^{-2}x = x^{-1}$$

or $R(x) = 1/x$. For any positive value of x, revenue will rise if x falls. Since costs also fall as x falls, then starting from any positive output level, this firm can increase profit by reducing output, provided that the firm does not reduce output to zero. This result follows because elasticity less than one implies that a reduction in output is accompanied by a larger increase, in percentage terms, of price. Therefore revenue rises when output is reduced as long as $x > 0$. As in the case of unit elasticity of demand, there is no positive value of output which leads to maximum profit. ■

S6.3 Average and Marginal Functions Revisited

In models of the firm, the relation between average and marginal product, or average and marginal cost, is a subject of some interest, particularly the proposition that the curves intersect at an extreme value of the average function (given that this occurs at a positive value of x). The marginal-product curve cuts the average-product curve at its maximum; the marginal-cost curve cuts the average-cost curve at its minimum. This is very easy to show. Let $f(x)$ denote the total-product or total-cost function, $a(x) \equiv f(x)/x$ the average-product or average-cost function, and, of course, $f'(x)$ the marginal function. Then

$$\begin{aligned} a'(x) &= \frac{1}{x^2}(xf'(x) - f(x)) \\ &= \frac{1}{x}\left(f'(x) - \frac{f(x)}{x}\right) \\ &= \frac{1}{x}(f'(x) - a(x)) \end{aligned}$$

It follows that when $a(x)$ is at an extreme value x^*

$$a'(x^*) = \frac{1}{x^*}(f'(x^*) - a(x^*)) = 0$$

implying that

$$f'(x^*) = a(x^*)$$

If $f(x)$ is strictly concave (as with the total-product function, at least over the "relevant range"), then $a(x)$ takes a maximum at x^*, while if $f(x)$ is strictly convex (as with the total-cost function), then $a(x)$ takes a minimum at x^*. Given some specific function for $f(x)$, we can always find the value of x^* at which $a(x)$ and $f'(x)$ are equal by using the above first-order condition.

Example S6.4 Suppose that we have a total-product function

$$f(x) = 10x + 12x^2 - x^3$$

The average-product function is

$$a(x) \equiv \frac{f(x)}{x} = 10 + 12x - x^2$$

and the marginal-product function is

$$f'(x) = 10 + 24x - 3x^2$$

At its maximum the average-product function's first derivative is zero:

$$a'(x^*) = 0 \Rightarrow 12 - 2x^* = 0 \Rightarrow x^* = 6$$

Notice that at $x^* = 6$ we have

$$a(x^*) = 10 + 12(6) - 6^2 = 46$$

$$f' = 10 + 24(6) - 3(6^2) = 46$$

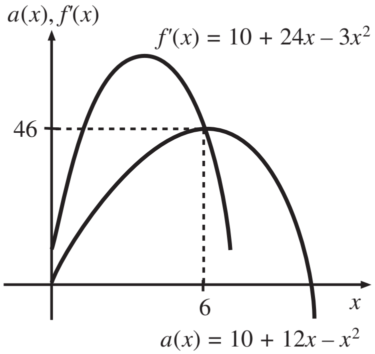

Figure S6.4 Graph of the result that shows that $a(x) = f(x)/x$ and $f'(x)$ have the same value where $a(x)$ achieves its extreme value

This result is illustrated in figure S6.4. Notice that $f''(x) = 24 - 6x < 0$ at $x^* = 6$, which indicates that the production function is concave at the point where $a(x)$ takes a maximum. ■

S6.4 The Labor-Managed Firm

Suppose that a monopoly is owned by its workers, who are paid in the following way: the firm's total revenue is shared equally among all the workers. Let $R(x)$ denote the firm's revenue function, $x = f(L)$ its production function, and $s = R/L$ the payment each worker receives. Then it seems reasonable to assume that this firm will choose its labor force so as to maximize s, and we have the problem

$$\max \frac{R[f(L)]}{L}$$

giving first-order condition

$$\frac{1}{L^2}(LR'f' - R[f(L)]) = 0$$

This can be rewritten as

$$R'f' = \frac{R}{L} = s$$

so that the number of workers is set at the point at which the marginal and average revenue products of labor are equal. Assuming that the revenue and production functions are both strictly concave, this will be at the maximum point of the average-revenue product of labor function.

S6.5 Competitive Firm with a Cubic Cost Function

The usual U-shaped average- and marginal-cost curves of the economics textbooks can be generated by total cost functions of the form

$$C(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

which is a third-degree polynomial or cubic equation. Suppose that we have a firm selling into a competitive market at a given price $p = \$30$, and with the cost function

$$C(x) = 20 + 50x - 3x^2 + 0.1x^3$$

In this function, 20 is the *fixed-cost* component, since it does not involve a term in output x, and the remainder of the function gives *total-variable cost*. The firm's revenue function is $R(x) = 30x$, and so its profit is

$$\pi(x) = 30x - [20 + 50x - 3x^2 + 0.1x^3]$$

The condition for profit maximization yields

$$\pi'(x^*) = 30 - [50 - 6x + 0.3x^2] = 0$$

Since the term in square brackets is marginal cost, this gives the familiar "price = marginal cost" condition. Simplifying the equation gives

$$-0.3x^2 + 6x - 20 = 0$$

a quadratic equation. Solving this gives two roots:

$$x_1^* = 4.2, \quad x_2^* = 15.8$$

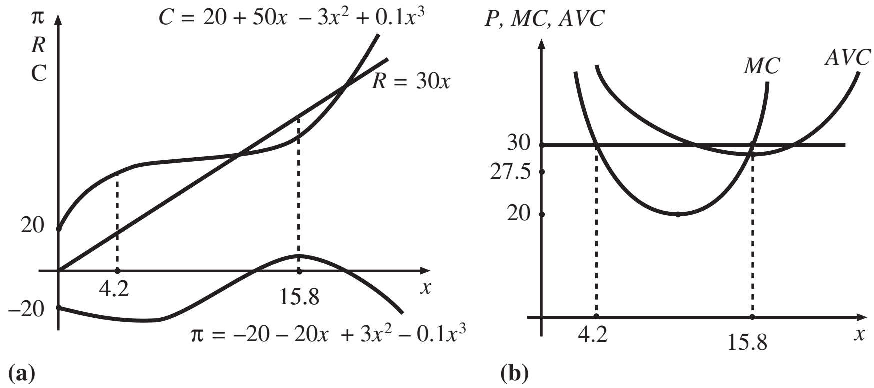

Figure S6.5 Competitive firm with cubic costs

Is this then a case where we have more than one local maximum, or is there something more involved? Figure S6.5 shows that only $x_2^* = 15.8$ is a local *maximum,* while $x_1^* = 4.2$, which, of course, is also a point at which price equals marginal cost, is in fact a *local minimum,* since profit is at its lowest (most negative) there. Thus, as we saw in the general discussion earlier (review figures 6.2 and 6.3), the condition $f'(x) = 0$ can characterize *both* maximum *and* minimum solutions.

For the remainder of this example, we focus on the maximum solution $x_2^* = 15.8$. At this point, the firm's total profit is

$$\begin{aligned}\pi(15.8) &= 30(15.8) - [20 + 50(15.8) - 3(15.8)^2 + 0.1(15.8)^3]\\ &= \$18.49\end{aligned}$$

and so production is profitable.

Suppose, however, that the firm's fixed cost had been not \$20 but \$40. Then an output of 15.8 is still the optimal solution, since this condition is unaffected by the value of the fixed cost, but the firm has a loss of \$1.51. Would it then be better to shut down and produce zero output, rather than produce 15.8 units at a loss? Table S6.1 gives the answer. It is clearly better to produce 15.8 units, since the loss is lower than if no output is produced.

Table S6.1 illustrates an important economic point that has validity far beyond this example. Note that the fixed costs are *unavoidable:* they are incurred regardless of whether the firm produces any output (in economic terminology we would say that the firm is "in the short run" since it has fixed costs). In deciding whether to produce, we can therefore ignore these fixed costs. We note that the firm's total revenue exceeds total variable costs, so an operating surplus (here \$38.49) helps offset its fixed cost. Clearly, as long as this operating surplus, revenue *minus* total variable cost, is positive, the firm should continue to produce. If this were negative,

Table S6.1 Shut down or keep producing?

Output	Revenue \$	Variable cost \$	Fixed cost \$	Loss \$
15.8	474	435.45	40	1.51
0	0	0	40	40

however, the firm should shut down, since otherwise it would be adding the loss it incurs in production to the fixed costs it has to pay anyway. If we denote the firm's total variable costs by the function $V(x)$, we can summarize by saying that the firm should produce the output $x^* > 0$ that satisfies its profit-maximizing condition as long as

$$R(x^*) \geq V(x^*)$$

Otherwise, it should set $x = 0$. Dividing through this inequality by output x^*, and noting that $R(x^*)/x^* = p$, while $V(x^*)/x^*$ is average variable cost, AVC, we can state the condition as

$$p \geq \text{AVC}$$

In figure S6.5(b) we see that this condition is satisfied at $x^* = 15.8$, since at that output AVC $= \$27.56 < p = \$30$.

S6.6 Short-Run Supply Function of a Competitive Firm

We continue to use this model of the competitive firm but now generalize slightly. Instead of taking a specific value for price, p, we leave it as a general parameter in the problem. The firm's profit function is now

$$\pi(x) = px - [20 + 50x - 3x^2 + 0.1x^3]$$

and the condition for profit maximization is

$$\pi'(x) = p - [50 - 6x + 0.3x^2] = 0$$

The firm's *supply function* gives its desired output x as a function of market price, p. This requires us to solve for x as a function of p. We can write the quadratic as

$$-0.3x^2 + 6x + (p - 50) = 0$$

and using the standard quadratic formula, we then have

$$x^* = \frac{-6 \pm \sqrt{36 - 4(-0.3)(p - 50)}}{(-0.6)}$$

Now we know from our earlier calculations that the profit *maximum* is at the higher of the two roots to this quadratic, and we have the resulting function

$$\begin{aligned} x^* &= [-6 - (1.2p - 24)^{1/2}]/-0.6 \\ &= 10 + 1.67(1.2p - 24)^{1/2} \end{aligned}$$

which gives x^* as a real number only for $p \geq 20$. (This corresponds to the rising part of the marginal cost curve—refer to figure S6.5(b).) We know, however, that for some range of prices above \$20, namely those that are less than AVC, the firm will produce zero output. In diagrammatic terms, the firm's supply curve is that portion of its marginal-cost curve that is above its AVC curve. To characterize this, we first need to solve for minimum AVC. The AVC function is given by

$$\frac{\text{TVC}}{x} = \frac{(50x - 3x^2 + 0.1x^3)}{x} = 50 - 3x + 0.1x^2$$

To minimize this function we set its derivative equal to zero:

$$-3 + 0.2x = 0$$

implying that AVC is minimized at $x = 15$. The resulting minimized value of AVC is

$$50 - 3(15) + 0.1(15)^2 = \$27.50$$

So we can say that at any price below \$27.50 the firm will shut down, while at a price at or above \$27.50 it will produce the output given by the first-order condition. This implies that its (short-run) supply function is

$$x = \begin{cases} 10 + 1.67(1.2p - 24)^{1/2} & \text{for } p \geq \$27.50 \\ 0 & \text{for } p < \$27.50 \end{cases}$$

S6.7 Competitive Firm with Cubic Costs Revisited

Return to the example of the competitive firm with a cubic-cost function. There we obtained the equation on profit-maximizing output

$$\pi'(x) = 30 - [50 - 6x + 0.3x^2] = 0$$

which we saw had two solutions; $x_1^* = 4.2$, $x_2^* = 15.8$. We chose between them by drawing a diagram. Instead, we can use the second-order condition. Now

$$\pi''(x) = 6 - 0.6x$$

So

$$\pi''(4.2) = 6 - 2.52 = 3.48 > 0$$

$$\pi''(15.8) = 6 - 9.48 = -3.48 < 0$$

Thus, from theorems 6.2 and 6.3, we can conclude that $x^* = 4.2$ yields a local minimum of the profit function and $x^* = 15.8$ yields a local maximum.

Example S6.5 **The Excise Tax That Maximizes Total Tax Revenue**

Find the excise tax, t, that will maximize total tax revenue in the following market:

$$D = 50 - 2p_B, \qquad S = -10 + p_S$$

where p_B is the price paid by consumers inclusive of tax and p_S is the price received by producers after the tax has been passed to the government. Check that your answer really delivers a maximum and not a minimum.

Solution

Clearly,

$$p_S = p_B - t$$

Therefore

$$D = 50 - 2p_B, \quad S = -10 + (p_B - t)$$

Equilibrium ($D = S$) implies that

$$50 - 2p_B = -10 + (p_B - t) \Rightarrow 3p_B = 60 + t$$

or

$$\hat{p}_B = \frac{60 + t}{3}$$

will be the equilibrium price. Equilibrium quantity is

$$\hat{D} = 50 - 2\left(\frac{60 + t}{3}\right) = \frac{30 - 2t}{3}$$

(You can check the result by substituting $\hat{p}_B$ into the supply function as well.) Thus total tax revenue is

$$T(t) = t\hat{D} = t\left(\frac{30 - 2t}{3}\right) = \frac{30t - 2t^2}{3}$$

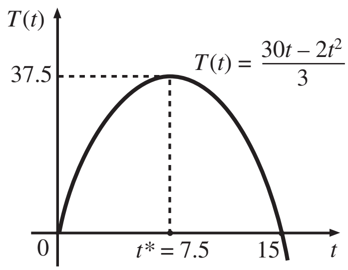

Figure S6.6 Graph of the tax revenue function for example S6.5

The value of t, $t = t^*$, that leads to maximum tax revenue, is found by setting $T'(t^*) = 0$. So

$$T'(t^*) = \frac{30 - 4t^*}{3} = 0 \Rightarrow t^* = 7.5$$

Notice that $T''(t) = -4/3 < 0$ for all t, implying that the function $T(t)$ is strictly concave everywhere and so $t^* = 7.5$ delivers a global maximum. This is illustrated in figure S6.6. Notice that $T(t) = 0$ at $t = 0$, for obvious reasons, and at $t = 15$, since once tax is this large, equilibrium price becomes

$$\hat{p}_B = \frac{60 + 15}{3} = 25$$

and demand is zero at this price. If nothing is purchased, tax revenue must be zero. The maximum tax revenue is

$$T(t^*) = \frac{30(7.5) - 2(7.5)^2}{3} = 37.5 \qquad \blacksquare$$

Chapter S7 Systems of Linear Equations

Contents

S7.1 Open-Economy IS-LM-BP Model

An important extension of the IS-LM model includes consideration of open economy issues: the trade in goods and services, and the international movement of capital. The two key endogenous variables of the closed economy IS-LM model, the level of GDP (Y) and the interest rage (R), are now potentially influenced by world markets. As we add this new dimension to the IS-LM model, we have to revisit assumptions about the parameters of the model and the variables we wish to have determined by it.

We can extend the linear two-sector IS-LM model to allow for trade and balance of payments considerations. The first implication of doing so is that net demand on domestic output comes from exports, less the value of imports. This is *net exports,* denoted by X. Net exports will be lower the higher is domestic income because an increase in income increases imports but has no effect on exports. In addition net exports will depend on the exchange rate, E, defined as the price of a foreign currency in terms of domestic currency. If the domestic economy is the United States and the foreign economy is Japan, E has the dimension \$ U.S./yen. An increase in E, so defined, makes U.S. goods relatively cheaper for Japanese consumers to buy, because each yen will buy more U.S. dollars. Simultaneously an increase in E makes Japanese goods relatively more expensive for U.S. residents to buy, since each dollar buys fewer yen. Thus an increase in E increases net exports to Japan. The IS side of the story is therefore determined by

$$
\begin{aligned}
C &= a + bY \\
I &= e - lR \\
X &= \bar{X} - mY + \alpha E
\end{aligned}
$$

representing the consumption function, investment function, and net export function respectively. Here $\bar{X}$ represents some exogenous level of net exports. Note that we are assuming no government sector here.

The LM sector is assumed to be unchanged, and we have

$$R = \frac{kY - \bar{M}}{h}$$

Finally, we will require that a full equilibrium involves a zero balance of payments. The balance of payments, B, is the sum of the capital account surplus (the net receipts from the sale of domestic assets to overseas residents) and net exports. A simple linear formulation is therefore

$$B = \bar{X} - mY + p(R - R^w) + \alpha E$$

The term $p(R - R^w)$ represents the influence of the capital account. The capital account surplus increases if the domestic interest rate is above the world rate. If the domestic interest rate is less than the world rate, then individuals can earn a higher rate of return on their assets overseas and a net outflow of capital will ensue. The exchange rate would also normally influence capital movements, but for simplicity we restrict the influence of the exchange rate to net exports. This makes the model rather special. Setting $Y = C + I + X$ and setting $B = 0$ give the IS curve and the BP (balance of payments equilibrium) curves, respectively. These are

$$R = \frac{a + e + \bar{X}}{l} - \frac{(1 - b + m)}{l}Y + \frac{\alpha}{l}E$$

$$R = \frac{mY - \bar{X} - \alpha E}{p} + R^w$$

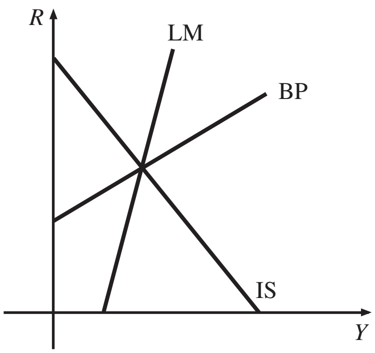

Figure S7.1 IS-LM-BP equilibrium

The IS, LM, and BP curves are shown in figure S7.1. Since both LM and BP curves are positively sloped, we need to make some assumption about the relative steepness of the lines. If we assume that capital is mobile but not perfectly mobile (see example 7.9), then BP may be less steep than LM. The slope of LM is

$$\left.\frac{dR}{dY}\right|_{\text{LM}} = \frac{k}{h}$$

and the slope of BP is

$$\left.\frac{dR}{dY}\right|_{\text{BP}} = \frac{m}{p}$$

We assume the latter is smaller than the former. The parameter p tells us how sensitive BP is to the interest rate differential. The larger is p the bigger are the induced capital movements when the interest rate differential changes. As p becomes very large, the slope of the BP curve becomes zero, which is essentially the situation we had in example 7.9.

How the equilibrium is determined now depends on the assumption to be made about the exchange rate. We will assume that the exchange rate is *flexible,* so that it becomes the third variable (in addition to output and interest rate) to be determined. We have three independent equations in three unknowns, Y, R, and E. The three reduced forms are

$$Y = \frac{h(a+e) + \bar{M}(l+p) + hpR^w}{(l+p)k + h(1-b)}$$

$$R = \frac{k(a+e) + kpR^w - (1-b)\bar{M}}{(l+p)k + h(1-b)}$$

$$E = \frac{(a+e)(hm-pk)}{\alpha[k(l+p)+h(1-b)]} + \bar{M}\left\{\frac{p(1-b+m)+lm}{\alpha[k(l+p)+h(1-b)]}\right\} - \frac{\bar{X}}{\alpha} + pR^w\left\{\frac{kl+h(1-b+m)}{\alpha[k(l+p)+h(1-b)]}\right\}$$

Example S7.1 Open-Economy Equilibrium with a Flexible Exchange Rate

You are given the following information about an open economy, where all the notation has been defined earlier:

$$C = 210 + 0.8Y$$
$$I = 20 - R$$
$$X = 200 - 0.3Y + 0.5E$$

$$L = 0.5Y - R$$
$$M = 480$$

$$B = 200 - 0.3Y + (R - 10) + 0.5E$$

Show that the BP curve is less steep than the LM curve in this case (the situation in figure S7.1). Solve for the equilibrium Y, R, and E.

Solution

In terms of the earlier notation, $k = 0.5$, $h = 1$, so the LM curve has a slope of 0.5. With $m = 0.3$ and $p = 1$, the BP curve has an absolute slope of 0.3.

We can solve for the equilibrium by using the equilibrium conditions

$$\begin{aligned} Y &= C + I + X \\ M &= L \\ B &= 0 \end{aligned}$$

Solving by substitution gives

$$R = 20\%, \quad Y = 1000, \quad E = 180$$

Notice that substituting the parameter values into the general reduced forms derived earlier produces approximately the same answers. Any discrepancies are due to rounding errors. ■

A special case of the system in equation (7.14) arises when all of the constants, b_i, are equal to zero.

S7.2 Gauss-Jordan Elimination

Some of the steps in producing the reduced row-echelon form of matrix from which the solution to the system is derived seem somewhat arbitrary or haphazard. The choice of procedure at each step seems to anticipate the future steps required to solve the system and it seems that some intuition and subjective judgment are at work. **Gauss-Jordan elimination** is a systematic procedure or *algorithm* which always leads to a matrix in reduced row-echelon form. The method involves the systematic interchange of rows and the application of row operations. Rather than state the algorithm generally, it is better illustrated by an example:

Example S7.2 **Open-Economy Equilibrium with a Flexible Exchange Rate**

Consider the system

$$\begin{array}{rcrcrcrcr} & & x_2 & + & 2x_3 & - & 4x_4 & = & -9 \\ 2x_1 & - & x_2 & & & + & x_4 & = & 6 \\ x_1 & + & x_2 & + & x_3 & + & x_4 & = & 14 \\ & & & - & x_3 & + & x_4 & = & 1 \end{array}$$

which may be written in matrix form as

$$\begin{bmatrix} 0 & 1 & 2 & -4 & -9 \\ 2 & -1 & 0 & 1 & 6 \\ 1 & 1 & 1 & 1 & 14 \\ 0 & 0 & -1 & 1 & 1 \end{bmatrix}$$

Solution

Step 1 Identify the first column to contain *any* nonzero elements. In this example we identify column 1.

Step 2 Interchange rows so that a 1 appears at the top of the identified column. If no row already contains a 1 in the identified column, but contains some other number α, then multiply the entire row by $1/\alpha$ to create a 1, then interchange rows so that the 1 appears at the top of the identified column. In this example we interchange the first and third rows to obtain

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 14 \\ 2 & -1 & 0 & 1 & 6 \\ 0 & 1 & 2 & -4 & -9 \\ 0 & 0 & -1 & 1 & 1 \end{bmatrix}$$

Step 3 Multiply the top row by the appropriate factor and subtract it from each row below to obtain zeros in the first column in all rows except the top. In this case we only need to subtract from the second row, twice the top row to give

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 14 \\ 0 & -3 & -2 & -1 & -22 \\ 0 & 1 & 2 & -4 & -9 \\ 0 & 0 & -1 & 1 & 1 \end{bmatrix}$$

Step 4 Ignore the top row for now. Identify in the remaining rows the first column to contain *any* nonzero element, and repeat steps 2 and 3. In our example we identify the second column as the leftmost column to contain any nonzero element, and we may interchange the second row with the third row to obtain a 1 at the top of the identified column:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 14 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 1 & 2 & -4 & -9 \\ 0 & -3 & -2 & -1 & -22 \\ 0 & 0 & -1 & 1 & 1 \end{bmatrix}$$

Operations are now on the lower submatrix. We now multiply the new first row by 3 and add to the second row to produce a zero below the 1. The new submatrix is

$$\begin{bmatrix} 0 & 1 & 2 & -4 & -9 \\ 0 & 0 & 4 & -13 & -49 \\ 0 & 0 & -1 & 1 & 1 \end{bmatrix}$$

Step 4 may now be repeated on this submatrix. The top row is ignored and joins the previously eliminated top row:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 14 \\ 0 & 1 & 2 & -4 & -9 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 4 & -13 & -49 \\ 0 & 0 & -1 & 1 & 1 \end{bmatrix}$$

The first nonzero column in the lower submatrix is the third column and we can multiply the second row by -1 and interchange with the top row to give

$$\begin{bmatrix} 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 4 & -13 & -49 \end{bmatrix}$$

Multiplying the first row by 4 and subtracting from the second row gives

$$\begin{bmatrix} 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & -9 & -45 \end{bmatrix}$$

The new top row can be ignored for now and joins the others, leaving just the last row to be divided by -9. Bringing all the previously ignored rows together with the new last row gives

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 14 \\ 0 & 1 & 2 & -4 & -9 \\ 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 1 & 5 \end{bmatrix}$$

Step 5 Starting with the last row, add appropriate multiples of each row to the rows above until the reduced row-echelon form is revealed. In the example we add the last row to the third row to obtain a zero *above* the 1 in the last row. We then add 4 times the last row to the second row to give

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 14 \\ 0 & 1 & 2 & 0 & 11 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 5 \end{bmatrix}$$

We then subtract 2 times the third row from the second row to give

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 14 \\ 0 & 1 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 5 \end{bmatrix}$$

Finally, we subtract the sum of the second, third, and fourth rows from the first row

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 5 \end{bmatrix}$$

to give us the reduced row-echelon form. The solution is therefore $x_1 = 2$, $x_2 = 3$, $x_3 = 4$, and $x_4 = 5$. ■

The Gauss-Jordan elimination technique is composed of two main parts. Steps 1 through 4 produce the zeros in the lower left-hand part of the matrix, while step 5 produces the zeros in the top right-hand part.

Chapter S8 Matrices

Contents

S8.1 Migration

The transition matrix summarizes information that characterizes the transition between "states." Examples of such states may be social classes, income groups, or geographical regions. We will examine the transition matrix that applies to population migration between regions. Workers may move or stay, depending on the economic conditions they face. For instance, for those with steady employment, staying put at the location of their employment is very sensible, but for those workers out of work, moving to another region may be worthwhile. Suppose that we have a country divided into three regions: 1, 2, and 3. Then the proportions of the populations of these regions that stay put or move to another region are given in terms of the transition matrix

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

The entry p_{ij} in this 3×3 matrix denotes the proportion of the population of region j that moves to region i, $j = 1, 2, 3$ and $i = 1, 2, 3$. For instance, if 80% of the population of region 1 stays put, 10% moves to region 2, and 10% moves to region 3, the corresponding entries of P are $p_{11} = 0.8$, $p_{21} = 0.1$, and $p_{31} = 0.1$. Similarly, if 70% of the population of region 2 stays, 15% moves to region 1, and 15% to region 3, we have $p_{12} = 0.15$, $p_{22} = 0.7$, and $p_{32} = 0.15$. Finally, if 90% of region 3's population stays, 5% moves to region 1, and 5% to region 2, we have $p_{13} = 0.05$, $p_{23} = 0.05$, and $p_{33} = 0.9$. Then the matrix P is given as

$$P = \begin{bmatrix} 0.80 & 0.15 & 0.05 \\ 0.10 & 0.70 & 0.05 \\ 0.10 & 0.15 & 0.90 \end{bmatrix} \qquad \blacksquare$$

The transition matrix, as we will see later in the chapter, can be used to evaluate the population movements between regions over time.

These examples lead to the general definition of a matrix.

S8.2 Profit for a Multiproduct Firm

Another application for matrix multiplication is in calculating the profit for a multiproduct firm.

Example S8.1 Suppose that a firm produces three types of output, using two types of input. Its output quantities are given by the column vector

$$\mathbf{q} = \begin{bmatrix} 15{,}000 \\ 27{,}000 \\ 13{,}000 \end{bmatrix}$$

and the unit prices of these are given by the row vector $\mathbf{p} = [10 \quad 12 \quad 5]$. The amounts of inputs it uses are given by the column vector

$$\mathbf{z} = \begin{bmatrix} 11{,}000 \\ 30{,}000 \end{bmatrix}$$

and the input prices by the row vector $\mathbf{w} = [20 \quad 8]$. The firm's profit is given by

$$\begin{aligned} \Pi &= \mathbf{pq} - \mathbf{wz} \\ &= [10 \quad 12 \quad 5] \begin{bmatrix} 15{,}000 \\ 27{,}000 \\ 13{,}000 \end{bmatrix} - [20 \quad 8] \begin{bmatrix} 11{,}000 \\ 30{,}000 \end{bmatrix} \\ &= (150{,}000 + 324{,}000 + 65{,}000) - (220{,}000 + 240{,}000) = 79{,}000 \end{aligned}$$ ■

Chapter S9 Determinants and the Inverse Matrix

Contents

S9.1 Gauss-Jordan Elimination and the Inverse Matrix

The method for computing the inverse matrix that we will describe in this section is based on the application of the so-called elementary row operations that we saw in chapter 7 as a way of obtaining the solution of a system of simultaneous equations. A nonsingular matrix A of order n can be reduced to I_n by a series of elementary row or column operations defined below.

Definition S9.1

An **elementary row operation** involves any of the following three cases:

1. Interchanging any two rows
2. Adding a multiple λ of one row to another
3. Multiplying any row of a matrix by a scalar $\lambda \neq 0$

Similarly we can define an elementary column operation as follows:

1. Interchanging any two columns
2. Adding a multiple λ of one column to another
3. Multiplying any column of a matrix by a scalar $\lambda \neq 0$

Definition S9.2

A matrix obtained from the identity matrix I_n by means of an elementary row or column operation is called an **elementary matrix.**

Note that we only use row or column operations; we *should not* mix the two. Without any loss of generality we will concentrate only on row operations. Since there are three types of elementary operations, there are three types of elementary matrices, examples of which are given below.

To illustrate the different elementary matrix operations, we consider the following matrix:

$$E_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

E_1 is obtained from I_7 by interchanging rows 2 and 6.

$$E_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

E_2 is obtained from I_7 by multiplying row 2 times λ and adding the product to row 6.

$$E_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

E_3 is obtained from I_7 by multiplying the second row by $\lambda \neq 0$.

Performing a row operation of a given type on a matrix A is the same as premultiplying the matrix by an appropriate elementary matrix. We can reduce any matrix A with a nonzero determinant to I_n by an appropriate sequence of elementary row operations. In other words,

$$E_s E_{s-1} \cdots E_1 A = I$$

where E_s denotes the elementary matrix associated with the sth operation.

Example S9.1 Reduce matrix A to I_3 by an appropriate sequence of row operations:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -1 \\ 1 & 2 & 1 \end{bmatrix}$$

Solution

First subtract row 3 from row 1 to obtain

$$\begin{bmatrix} 0 & 0 & 2 \\ 0 & 1 & -1 \\ 1 & 2 & 1 \end{bmatrix}$$

Multiply row 2 by -2 and add the product to row 3 to obtain

$$\begin{bmatrix} 0 & 0 & 2 \\ 0 & 1 & -1 \\ 1 & 0 & 3 \end{bmatrix}$$

Next, multiply row 1 by $(-3/2)$:

$$\begin{bmatrix} 0 & 0 & -3 \\ 0 & 1 & -1 \\ 1 & 0 & 3 \end{bmatrix}$$

Now, add row 3 to row 1:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & 3 \end{bmatrix}$$

Subtract row 1 from row 3:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 3 \end{bmatrix}$$

Multiply row 3 by $1/3$:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

Finally, add row 3 to row 2:

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

■

In this example we could have followed other sequences of elementary row operations to arrive at the same result.

If the *same* sequence of elementary row operations is applied to I_n, the result will be the inverse matrix A^{-1}. We can see this from the fact that

$$E_s E_{s-1} \cdots E_1 A = I_n$$

where $E_s, E_{s-1}, \cdots, E_1$ stand for the elementary matrices associated with the corresponding s elementary operations. Then

$$A = E_1^{-1} E_2^{-1} \cdots E_s^{-1}$$

and

$$A^{-1} = E_s E_{s-1} \cdots E_1$$

Example S9.2 Use the Gauss-Jordan elimination method to obtain the inverse of matrix A,

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -1 \\ 1 & 2 & 1 \end{bmatrix}$$

Solution

We will apply the same sequence of elementary operations to I_3 as were applied to A when we reduced it to I_3. So we start with

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -1 \\ 1 & 2 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

First, we subtract row 3 from row 1:

$$\begin{bmatrix} 0 & 0 & 2 \\ 0 & 1 & -1 \\ 1 & 2 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Now, multiply row 2 by -2 and add the product to row 3:

$$\begin{bmatrix} 0 & 0 & 2 \\ 0 & 1 & -1 \\ 1 & 0 & 3 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}$$

Multiply row 1 by $(-3/2)$:

$$\begin{bmatrix} 0 & 0 & -3 \\ 0 & 1 & -1 \\ 1 & 0 & 3 \end{bmatrix}, \quad \begin{bmatrix} -3/2 & 0 & 3/2 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}$$

Add row 3 to row 1:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & 3 \end{bmatrix}, \quad \begin{bmatrix} -3/2 & -2 & 5/2 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}$$

Subtract row 1 from row 3:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 3 \end{bmatrix}, \quad \begin{bmatrix} -3/2 & -2 & 5/2 \\ 0 & 1 & 0 \\ 3/2 & 0 & -3/2 \end{bmatrix}$$

Multiply row 3 by $(1/3)$:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} -3/2 & -2 & 5/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & -1/2 \end{bmatrix}$$

Now, add row 3 to row 2:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} -3/2 & -2 & 5/2 \\ 1/2 & 1 & -1/2 \\ 1/2 & 0 & -1/2 \end{bmatrix}$$

These matrices are I_3 and A^{-1}, respectively. Note that A^{-1} was obtained earlier through the computation of the appropriate adjoint matrix in example 9.15. ■

Chapter S10 Advanced Linear Algebra

Contents

S10.1 The Statistical Distribution of Quadratic Forms

In statistical distribution theory the multivariate normal distribution is the most widely used distribution and constitutes the building block on which other commonly used distributions are constructed. Let $\mathbf{z}$ be a vector of random variables of dimension $n \times 1$, each with zero mean and variance unity. Then we say that

$$\mathbf{z} \sim N(\mathbf{0}, I)$$

which is read: $\mathbf{z}$ is distributed as a normal variable with mean the zero vector and variance-covariance matrix the identity matrix of order n. The distribution of $\mathbf{z}$ is characterized entirely in this case by its mean and its variance. The mean is defined as $E(\mathbf{z})$ and the variance of $\mathbf{z}$ is given by the variance-covariance matrix defined as

$$\text{var}(\mathbf{z}) = E(\mathbf{z} - E(\mathbf{z}))(\mathbf{z} - E(\mathbf{z}))^T$$

This matrix is a symmetric matrix of dimension $n \times n$. We can write the mean and variance of $\mathbf{z}$ explicitly as

$$E(\mathbf{z}) = E\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} E(z_1) \\ E(z_2) \\ \vdots \\ E(z_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}$$

and

$$\operatorname{var}(\mathbf{z}) = E\left(\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} [z_1 z_2 \cdots z_n]\right) = E\begin{bmatrix} z_1^2 & z_1 z_2 & \cdots & z_1 z_n \\ z_2 z_1 & z_2^2 & \cdots & z_2 z_n \\ \vdots & \vdots & \ddots & \vdots \\ z_n z_1 & z_2 z_n & \cdots & z_n^2 \end{bmatrix}$$

$$= \begin{bmatrix} E\left(z_1^2\right) & E(z_1 z_2) & \cdots & E(z_1 z_n) \\ E(z_2 z_1) & E(z^2) & \cdots & E(z_2 z_n) \\ \vdots & \vdots & \ddots & \vdots \\ E(z_n z_1) & E(z_n z_2) & \cdots & E\left(z_n^2\right) \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I_n$$

Note that $E(z_i z_j)$ are the correlations among the z_is and the z_js, $(i, j = 1, \ldots, n)$, which are assumed to be zero.

The sum of the squares of these independently and identically distributed z_is defines a common distribution in econometrics, the χ^2 distribution. ■

Example S10.1 The Classical Least-Squares Model

In this example, we will discuss the classical linear model and its basic mathematical and statistical structure. We will highlight the results from linear algebra and statistical distribution theory that we use.

Let us describe the

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}$$

where $\mathbf{y}$ is an $(n \times 1)$ vector of observations on the dependent variable, X is an $(n \times k)$ matrix of observations on k independent variables, $\boldsymbol{\beta}$ is a $(k \times 1)$ vector of parameters, and $\mathbf{u}$ is a vector which is $(n \times 1)$ and whose components are random errors or disturbances that are unobservable. A model where the dependent variable is expressed in terms of independent variables is known as a *regression model*. We assume the following:

1. The u_is are random variables with the properties

$$E[\mathbf{u}] = E\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} E(u_1) \\ E(u_2) \\ \vdots \\ E(u_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}$$

and

$$\mathrm{var}(\mathbf{u}) = E\{(\mathbf{u} - E(\mathbf{u}))(\mathbf{u} - E(\mathbf{u}))^T\} = E(\mathbf{u}\mathbf{u}^T)$$

since $E(\mathbf{u}) = 0$. More explicitly

$$\begin{aligned}
\mathrm{var}(\mathbf{u}) &= E\left[\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} [u_1 u_2 \cdots u_n]\right] \\
&= E\begin{bmatrix} u_1^2 & u_1u_2 & \cdots & u_1u_n \\ u_2u_1 & u_2^2 & \cdots & u_2u_n \\ \vdots & \vdots & \ddots & \vdots \\ u_nu_1 & u_nu_2 & \cdots & u_n^2 \end{bmatrix} \\
&= \begin{bmatrix} E\left(u_1^2\right) & E(u_1u_2) & \cdots & E(u_1u_n) \\ E(u_2u_1) & E\left(u_2^2\right) & \cdots & E(u_2u_n) \\ \vdots & \vdots & \ddots & \vdots \\ E(u_nu_1) & E(u_nu_2) & \cdots & E\left(u_n^2\right) \end{bmatrix} \\
&= \sigma^2 I_n
\end{aligned}$$

where I_n is the identity matrix of order n. The assumption above about the errors simply states that they are pairwise uncorrelated, since $E(u_iu_j) = 0$, for all $i \neq j$. They also have the same variances. If we further assume that the joint distribution of these n errors is normal, then they will be independent, since for the case of normality, lack of correlation implies independence, and vice versa. Then we say that the u_is are *independently and identically distributed,* or *i.i.d.,* and that they are *spherical*.

2. The Xs are fixed, real numbers. In other words, in contrast to the random behavior of the u_is the Xs do not add to the randomness that is transmitted to the the ys. Hence the stochastic nature of the model is entirely due to the randomness of the errors. We will also assume that the matrix X^TX will be nonsingular.

Least-squares estimation involves minimizing the sum of squared deviations of the predicted from the actual series. The solution that minimizes these squared deviations leads to the best choice of estimator. That is, for each choice of estimator, there is a corresponding predicted or estimated or fitted value of the dependent value and we choose the estimator that makes these fitted or predicted values mimic the actual data as well as possible. The problem is to minimize the objective function given by

$$\begin{aligned}\sum_{i=1}^{n} e_i^2 = \mathbf{e}^T\mathbf{e} &= (\mathbf{y} - X\mathbf{b})^T(\mathbf{y} - X\mathbf{b}) \\ &= \mathbf{y}^T\mathbf{y} - \mathbf{b}^T X^T\mathbf{y} - \mathbf{y}^T X\mathbf{b} + \mathbf{b}^T X^T X\mathbf{b}\end{aligned}$$

Note that the objective function is defined in terms of observable quantities, the e_i, which are the *residuals* from the estimation. The residuals are to be distinguished from the errors, the u_i, that are unobservables. We use $\mathbf{b}$ to denote the estimator of β, which is, of course, unknown. Different choices of $\mathbf{b}$ lead to different values of the objective function, since the regression residuals will be different. We then try to choose that value of $\mathbf{b}$ that minimizes the above objective function. Since this objective function is a scalar, all of its components are also scalars. Hence $\mathbf{b}^T X^T\mathbf{y} = \mathbf{y}^T X\mathbf{b}$, since taking the transpose of a scalar leaves it unchanged. That is, $\mathbf{b}^T X^T\mathbf{y} = (\mathbf{y}^T X\mathbf{b})^T$. The objective function then becomes

$$\mathbf{e}^T\mathbf{e} = \mathbf{y}^T\mathbf{y} - 2\mathbf{b}^T X^T\mathbf{y} + \mathbf{b}^T X^T X\mathbf{b}$$

In chapter 12 we will investigate the solution to the above minimization problem. The optimal choice of $\mathbf{b}$ is known as the *ordinary least-squares* (OLS) *estimator* of β. ■

Example S10.2 The Generalized Least-Squares Transformation

Suppose that the linear regression model $\mathbf{y} = X\beta + \mathbf{u}$ has errors that are nonspherical. In this case $\mathbf{u} \sim N(\mathbf{0}, \sigma^2\Omega)$, where Ω is a positive definite matrix. However, $\Omega \neq I$, and it may have a nonconstant main diagonal and/or possibly nonzero diagonal elements. The classical least-squares model of example 10.24 has certain desirable statistical properties that are partly the result of the i.i.d. (spherical) structure of the errors. Given that the errors in the present model are $\mathbf{u} \sim N(\mathbf{0}, \sigma^2\Omega)$, we need to transform them so that they will be spherical. In other words, we want to find a transformation T such that the variance-covariance matrix of the transformed errors $T\mathbf{u}$ is $\sigma^2 I$. We take T to be an $n \times n$ fixed matrix. The variance of $T\mathbf{u}$ is given below:

$$\text{var}(T\mathbf{u}) = E(T\mathbf{u} - E(T\mathbf{u}))(T\mathbf{u} - E(T\mathbf{u}))^T$$

Since $E(T\mathbf{u}) = TE(\mathbf{u}) = \mathbf{0}$, we get

$$\begin{aligned}\text{var}(T\mathbf{u}) &= E(T\mathbf{u}\mathbf{u}^T T^T) \\ &= TE(\mathbf{u}\mathbf{u}^T)T^T \\ &= \sigma^2 T\Omega T^T\end{aligned}$$

If we choose T such that $T\Omega T^T = I$, then transforming the model and applying least squares to the transformed model will bring us back to the environment where least squares are optimal. Below we will demonstate that such a transformation exists. By assumption, we have that Ω is a positive definite symmetric matrix. Then we can use theorem 10.6 to write

$$Q^T \Omega Q = \Lambda$$

where Q is the orthogonal matrix of eigenvectors and Λ is the diagonal matrix of eigenvalues. Then, by pre- and postmultiplying the equation above by Q and Q^T, respectively, we get

$$\Omega = Q\Lambda Q^T$$

since $QQ^T = I$. We can take the square root of the eigenvalues because they are all positive. That leads us to

$$\Omega = Q\Lambda^{1/2}\Lambda^{1/2}Q^T = PP^T$$

where P is a positive definite matrix defined as $Q\Lambda^{1/2}$. By choosing $T = P^{-1}$, we obtain

$$\begin{aligned} T\Omega T^T &= P^{-1}(PP^T)(P^{-1})^T \\ &= P^{-1}PP^T(P^T)^{-1} = I \end{aligned}$$

since we can interchange inversion and transposition. Having found the appropriate transformation T, we can apply it to the model as a whole. We can write the transformed model as

$$T\mathbf{y} = TX\beta + T\mathbf{u}$$

Then the objective function to be minimized becomes

$$\begin{aligned} \mathbf{e}_*^T\mathbf{e}_* &= (T\mathbf{y} - TX\mathbf{b}_*)^T(T\mathbf{y} - TX\mathbf{b}_*) \\ &= (\mathbf{y} - X\mathbf{b}_*)^T T^T T(\mathbf{y} - X\mathbf{b}_*) \end{aligned}$$

Again the solution involves appropriately choosing $\mathbf{b}_*$ to minimize $\mathbf{e}_*^T\mathbf{e}_*$. The optimal choice of $\mathbf{b}_*$ is known as the *generalized least-squares* (GLS) *estimator* of β. ■

Chapter S11 Calculus for Functions of *n*-Variables

Contents

S11.1 More Properties of Homogeneous Functions' Homotheticity

The properties of homogeneous functions that are presented in the text are those that are especially important for use in economic models, notably for understanding models that include use of production functions. There are many other properties of homogeneous functions that are useful in economics and some of these are developed here. We also explore in this section of supplementary material how the related (and broader) class of homothetic functions can be generated from homogeneous functions.

A general property of homogeneous functions is illustrated by the graph of the level curves of the function $y = x_1^{1/2} x_2^{1/2}$ in figure 11.26. Starting from different input bundles on a given isoquant, we can see that if we scale each input by the same factor for each input bundle, we generate new input bundles that also lie on a single isoquant. Thus, upon drawing rays from the origin to illustrate the effect of multiplying different input bundles by the same scale factor, we see that the isoquants are radial expansions and contractions of each other. This is a general property of all homogeneous functions. We will state this formally as a theorem:

Theorem S11.1 Suppose that $y = f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}_+^n$ is a homogeneous function. If $\mathbf{x}^0$ and $\bar{\mathbf{x}}$ are any two points on the same level curve of the function f and we multiply each of these points by the same factor s to get points $s\mathbf{x}^0$ and $s\bar{\mathbf{x}}$ respectively, then both of these points will also lie on a single-level curve.

Proof

Since $\mathbf{x}^0$ and $\bar{\mathbf{x}}$ lie on the same level curve f, it follows that

$$f(\mathbf{x}^0) = f(\bar{\mathbf{x}}) = y^0$$

Letting k be the degree of homogeneity of f, we get

$$f(s\mathbf{x}^0) = s^k f(\mathbf{x}^0) = s^k y^0$$

and

$$f(s\bar{\mathbf{x}}) = s^k f(\bar{\mathbf{x}}) = s^k y^0$$

and so

$$f(s\mathbf{x}^0) = f(s\bar{\mathbf{x}}) = s^k y^0$$

That is, $s\bar{\mathbf{x}}$ and $s\mathbf{x}^0$ lie on the same level curve, with $y = s^k y^0$. ■

Theorem S11.1 can clearly be extended to include any number of points on the *initial* isoquant $f(x_1, x_2) = y^0$, as illustrated in Figure 11.26 for the Cobb-Douglas production function $y = x_1^{1/2} x_2^{1/2}$. An interesting consequence of theorems 11.14 and S11.1 is that if a production function is homogeneous, then whatever returns-to-scale property it possesses is independent of which input bundle

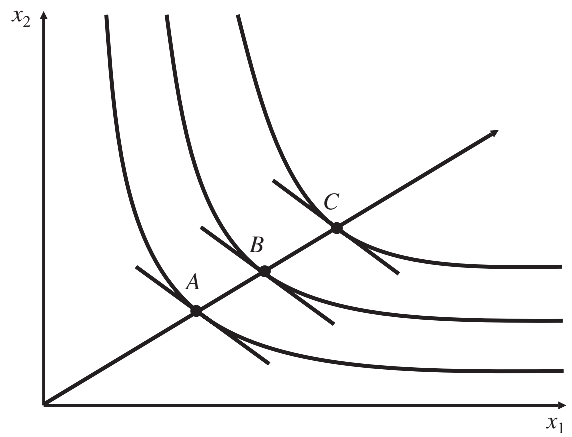

Figure S11.1 Slope of the isoquants. Note that the slope is the same at all points along a ray from the origin if the production function is homogeneous.

one begins with. For example, the production function $y = x_1^{1/2} x_2^{1/2}$ displays constant returns to scale no matter which input bundle one starts from. (Try doubling inputs beginning with bundle $(1, 9)$ and then $(4, 25)$.) This characteristic is not shared, for example, by the nonhomogeneous function examined in example 11.35.

Another interesting and useful property of homogeneous functions is that at points where a ray from the origin intersects the level curves of f, the slopes of the level curves are equal. For example, the slopes of the isoquants at points A, B, and C in figure S11.1 are equal if f is homogeneous. The following theorems are also useful.

Theorem S11.2 If f is a function which is homogeneous of degree k, then its first-order partial derivatives are homogeneous of degree $k - 1$.

Proof

Since f is homogeneous of degree k,

$$f(sx_1, sx_2, \ldots, sx_n) = s^k f(x_1, x_2, \ldots, x_n) \tag{S11.1}$$

Let $z_i = sx_i$ and then differentiate both sides with respect to x_i. Using the chain rule to evaluate the result on the left-hand side of equation (S11.1), we get

$$\frac{\partial f(\mathbf{z})}{\partial x_i} = \frac{\partial f}{\partial z_i} \frac{\partial z_i}{\partial x_i} = s^k \frac{\partial f}{\partial x_i}$$

But z_i is just another way of referring to the ith argument of f and so, using $\partial z_i / \partial x_i = s$, we have

$$f_i(\mathbf{z}) s = s^k f_i(\mathbf{x})$$

which upon dividing through by s gives the result

$$f_i(sx_1, sx_2, \ldots, sx_n) = s^{k-1} f_i(x_1, x_2, \ldots, x_n)$$

That is, the first-order partial derivatives are homogeneous of degree $k - 1$. ■

Theorem S11.3 If $y = f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^2_+$, is a production function that is homogeneous of degree 1, then its marginal products f_1 and f_2 depend only on the ratio of the input levels, x_1/x_2, and not on their absolute size.

Proof

From theorem S11.2 it follows that if the production function is homogeneous of degree one, then the marginal product functions are homogeneous of degree zero and so we have

$$f_1(sx_1, sx_2) = sf_1(x_1, x_2) = f_1(x_1, x_2) \tag{S11.2}$$

$$f_2(sx_1, sx_2) = sf_2(x_1, x_2) = f_2(x_1, x_2)$$

This relationship holds for any scalar value $s > 0$, and so it holds for the particular value $s = 1/x_2$. Therefore equations (S11.2) become

$$f_1\left(\frac{x_1}{x_2}, 1\right) = f_1(x_1, x_2)$$

$$f_2\left(\frac{x_1}{x_2}, 1\right) = f_2(x_1, x_2)$$

and so the value of the ratio x_1/x_2 determines completely the values of the marginal products. ■

Theorem S11.3 can be extended to the case of $y = f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n_+$, in which case one needs to select a specific x_i, $i = 1, 2, \ldots, n$ to form the input ratio. For example, if x_n is selected, then we can say the marginal products are determined completely by the values x_1/x_n, x_2/x_n, x_3/x_n, ..., x_{n-1}/x_n, and so multiplying all inputs by the same factor s will not change the values of the marginal products.

Example S11.1 Finding First-Order Partial Derivatives

Show that theorem S11.2 applies for the function $y = f(x_1, x_2) = x_1^{1/3}x_2^{1/4}$ by finding the first-order partials.

Solution

f is homogeneous of degree $k = 7/12$, since

$$f(sx_1, sx_2) = (sx_1)^{1/3}(sx_2)^{1/4} = s^{7/12}f(x_1, x_2)$$

The first-order partials are

$$f_1 = \frac{1}{3}x_1^{-2/3}x_2^{1/4}, \quad f_2 = \frac{1}{4}x_1^{1/3}x_2^{-3/4}$$

where the f_i are homogeneous of degree $k - 1 = -5/12$ (as implied by theorem S11.3). We can see this is so because

$$f_1(sx_1, sx_2) = \frac{1}{3}(sx_1)^{-2/3}(sx_2)^{1/4} = s^{-5/12} f_1(x_1, x_2)$$

and

$$f_2(sx_1, sx_2) = \frac{1}{4}(sx_1)^{1/3}(sx_2)^{-3/4} = s^{-5/12} f_2(x_1, x_2)$$

■

Example S11.2 Finding Marginal Product Functions

Show that theorem S11.3 applies for the production function $y = f(x_1, x_2) = x_1^{0.4}x_2^{0.6}$ by finding its marginal-product functions.

Solution

$$f_1 = 0.4x_1^{-0.6}x_2^{0.6} = 0.4\left(\frac{x_1}{x_2}\right)^{-0.6}$$

$$f_2 = 0.6x_1^{0.4}x_2^{-0.4} = 0.6\left(\frac{x_1}{x_2}\right)^{0.4}$$

Therefore the marginal products depend only on the ratio of the input values. ■

We now turn to the phenomenon illustrated in figure S11.1, namely the result that along a ray from the origin the slopes of isoquants are equal if f is homogeneous. Recall from section 11.3 that the negative of this slope, if f is a production function, is called the marginal rate of technical substitution (MRTS) and is equal to the ratio of the marginal products of f:

$$\text{MRTS} = -\frac{dx_2}{dx_1} = \frac{f_1}{f_2}$$

For the case with an arbitrary number of inputs, $y = f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n_+$, we have

$$\text{MRTS}_{k,l} = -\frac{dx_l}{dx_k} = \frac{f_k}{f_l}$$

where $\text{MRTS}_{k,l}$ is the MRTS between inputs k and l. We state and prove the result for the case of two inputs

Theorem S11.4 If $y = f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^2_+$ is a production function that is homogeneous of degree k and has continuous first-order partial derivatives, then along any ray from the origin the slope of all isoquants, or the MRTS, is equal.

Proof

Since the ratio x_1/x_2 is constant along any ray from the origin, we need to show that

$$\text{MRTS} = \frac{f_1}{f_2}$$

depends only on the ratio x_1/x_2. If f is homogeneous of degree k, then by theorem S11.2, f_1 and f_2 are homogeneous of degree $k-1$, and so

$$f_1(sx_1, sx_2) = s^{k-1} f_1(x_1, x_2)$$

and

$$f_2(sx_1, sx_2) = s^{k-1} f_2(x_1, x_2)$$

Choosing $s = 1/x_2$ gives us the result

$$\begin{aligned} \text{MRTS} &= \frac{f_1(x_1/x_2, 1)}{f_2(x_1/x_2, 1)} = \frac{(1/x_2)^{k-1} f_1(x_1, x_2)}{(1/x_2)^{k-1} f_2(x_1, x_2)} \\ &= \frac{f_1(x_1, x_2)}{f_2(x_1, x_2)} \end{aligned}$$

and so the ratio f_1/f_2 is completely determined by the ratio x_1/x_2. ■

It is important to note the difference between theorem S11.3 and theorem S11.4. If f is homogeneous of degree 1, then its marginal products, f_1 and f_2, are unchanged as one moves along a ray from the origin and so naturally the MRTS $= f_1/f_2$ is unchanged. However, if f is homogeneous of degree $k \neq 1$, the *ratio* of the marginal products, MRTS $= f_1/f_2$, is unchanged along a ray from the origin but the marginal products themselves are not constant. The following example illustrates:

Example S11.3 **Finding Marginal Rate of Technical Substitution**

Show that along any ray from the origin (i.e., x_1/x_2 constant) the slopes of the isoquants for the production function $y = x_1^{1/4}x_2^{1/2}$ are equal, although the marginal products do change.

Solution

The marginal-product functions are

$$f_1 = \frac{1}{4}x_1^{-3/4}x_2^{1/2}$$

$$f_2 = \frac{1}{2}x_1^{1/4}x_2^{-1/2}$$

which implies that

$$\text{MRTS} = \frac{f_1}{f_2} = \frac{\frac{1}{4}x_1^{-3/4}x_2^{1/2}}{\frac{1}{2}x_1^{1/4}x_2^{-1/2}} = \frac{1}{2}\frac{x_2}{x_1}$$

Thus the slopes of the isoquants are the same for x_1/x_2 constant. The marginal-product functions, however, are not completely determined by the ratio x_1/x_2. For example, if $x_1 = 5$, $x_2 = 8$ $(x_1/x_2 = 5/8)$, we get

$$f_1 = \frac{1}{4}5^{-3/4}8^{1/2} = 0.211$$

$$f_2 = \frac{1}{2}5^{1/4}8^{-1/2} = 0.264$$

while if $x_1 = 10$ and $x_2 = 16$ $(x_1/x_2 = 5/8)$, we get

$$f_1 = \frac{1}{4}10^{-3/4}16^{1/2} = 0.178$$

$$f_2 = \frac{1}{2}10^{1/4}16^{-1/2} = 0.222$$

■

S11.2 Homotheticity

The concept of returns to scale is not a meaningful one for utility functions. However, some of the properties of the level curves corresponding to homogeneous functions are useful in the context of consumer theory. In figure S11.2 we illustrate some level curves for some function $y = f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^2_+$, which are radial expansions and contractions of each other, as for a homogeneous function. However, from the function values attached to the level curves, it is clear that this function is not homogeneous (i.e., $f(2\mathbf{x}^0) = 2f(\mathbf{x}^0)$, implies homogeneity of degree 1, while $f(3\mathbf{x}^0) = 2.4f(\mathbf{x}^0)$ implies homogeneity of degree less than 1). It follows that the class of functions that satisfies the property that level curves are radial expansions and contractions of each other is larger than the set of functions that is homogeneous. It includes all homogeneous functions as well as all monotonic transformations of homogeneous functions. The latter are called **homothetic functions.**

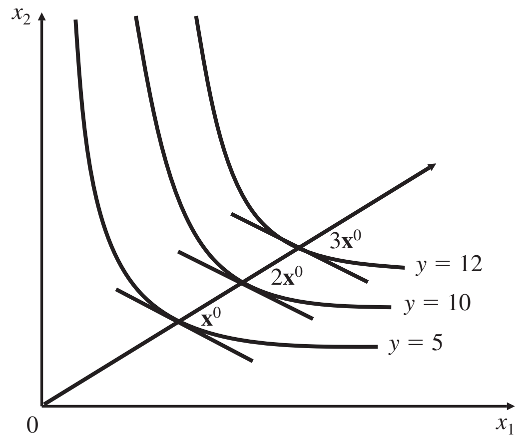

Figure S11.2 Level curves of a function that is homothetic but not homogeneous

Definition S11.1 A function is **homothetic** if it is a monotonic transformation of some homogeneous function.

Theorem S11.5 Let f be a function defined on $\mathbb{R}^2_+$. The function f is homothetic if and only if along any ray from the origin the slope of each level curve (i.e., the value of f_1/f_2) is constant.

This theorem extends to functions defined on $\mathbb{R}^n_+$ in the same way as does the equivalent result for homogeneous functions. That is, if a function f defined on $\mathbb{R}^n_+$ is homothetic, then the value of

$$-\frac{dx_l}{dx_k} = \frac{f_k}{f_l}$$

is unchanged if all x_i values are multiplied by the same factor.

Example S11.4 The function $f(x_1, x_2) = 1 + x_1^{1/2}x_2^{1/2}$ defined on $\mathbb{R}^2_+$ is not homogeneous but is homothetic.

Solution

We know already that $g(x_1, x_2) = x_1^{1/2}x_2^{1/2}$ is homogeneous, and since f is a monotonic transformation of g, it follows that f is homothetic. It is straightforward to show that f is not homogeneous. For instance, beginning with $(x_1^0, x_2^0) = (1, 1)$ we can see by comparing the values $f(1, 1) = 2$, $f(2, 2) = 3$, and $f(3, 3) = 4$ that there is no pattern $f(sx_1^0, sx_2^0) = s^k f(x_1^0, x_2^0)$. ■

Since we are only interested in the shape of the indifference curves implied by a utility function, and a utility function is unique only up to a positive, monotonic transformation, then only the requirement of homotheticity is relevant and not the additional properties implied by homogeneity.

PRACTICE EXERCISES

S11.1. Show that the MRTS for the Cobb-Douglas production function $f(x_1, x_2) = Ax_1^{\alpha}x_2^{\beta}$, $A > 0$, $\alpha, \beta > 0$ defined on $\mathbb{R}^2_{++}$ is constant along any ray from the origin. Also show that the marginal product functions change value along such a ray unless $\alpha + \beta = 1$.

S11.2. Show that the following functions are homothetic but not homogeneous:

(a) $f(x_1, x_2) = k + x_1^{1/2}x_2^{1/2}$ for $k \neq 0$ a constant

(b) $f(x_1, x_2) = e^{x_1^2 x_2}$

S11.3. Which of the following functions are (i) homogeneous, (ii) homothetic but not homogeneous, (iii) neither?

(a) $f(x_1, x_2, x_3) = x_1^{a_1}x_2^{a_2}x_3^{a_3}$ for $a_i > 0$ defined on $\mathbb{R}^3_{++}$

(b) $f(x_1, x_2, x_3) = 1 + x_1^{1/2} x_2^{1/4} x_3^{1/4}$ defined on $\mathbb{R}^3_{++}$

(c) $f(x_1, x_2, x_3) = x_1^{1/2} x_2^{1/3} + x_2^{3/2}$ defined on $\mathbb{R}^3_{++}$

Solutions

S11.1. $f_1 = \alpha A x_1^{\alpha-1} x_2^{\beta}$, $f_2 = \beta A x_1^{\alpha} x_2^{\beta-1}$. Along a ray from the origin $x_2 = kx_1$, so

$$\text{MRTS} = \frac{f_1}{f_2} = \frac{\alpha x_2}{\beta x_1} = \frac{\alpha k x_1}{\beta x_1} = \frac{\alpha k}{\beta}$$

which is a constant. Using $x_2 = kx_1$ in f_1 and f_2 shows that marginal products are independent of x_1 only if $\alpha + \beta = 1$. To see this, note that upon setting $x_2 = kx_1$ we obtain

$$f_1 = \alpha A k^{\beta} x_1^{\alpha+\beta-1}$$

and

$$f_2 = \beta A k^{\beta-1} x_1^{\alpha+\beta-1}$$

S11.2. **(a)** $x_1^{1/2} x_2^{1/2}$ is a homogeneous function

$$(sx_1)^{1/2}(sx_2)^{1/2} = s x_1^{1/2} x_2^{1/2}$$

(i.e., it is homogeneous of degree one). Thus, since f is a monotonic transformation of $x_1 x_2$, it follows that f is homothetic. To see that f is not homogeneous, note that $f(1, 1) = k + 1$, $f(2, 2) = k + 2$, and $f(4, 4) = k + 4$. If f is homogeneous, it must be the case that $f(2, 2) = 2^t f(1, 1)$ and $f(4, 4) = 2^t f(2, 2)$ for some value t. This requires that $(k + 2) = 2^t(k + 1)$ and $(k + 4) = 2^t(k + 2)$. Taking ratios, this means that $2^t = (k+2)/(k+1)$ and $2^t = (k+4)/(k+2)$, which in turn implies that $(k + 2)/(k + 1) = (k + 4)/(k + 2)$ or $(k + 2)(k + 2) = (k + 1)(k + 4)$. These results, however, are not compatible as

$$(k + 2)(k + 2) = k^2 + 4k + 4$$

$$(k + 1)(k + 4) = k^2 + 5k + 4$$

and it is not possible for

$$k^2 + 4k + 4 = k^2 + 5k + 4 \text{ or } 4k = 5k$$

unless $k = 0$. Thus f cannot be a homogeneous function.

(b) $x_1^2 x_2$ is a homogeneous function, since

$$(sx_1)^2(sx_2) = s^3 x_1^2 x_2$$

(i.e., it is homogeneous of degree three). Thus, since f is a monotonic transformation of $x_1 x_2$, it follows that f is homothetic. To see that f is not homogeneous, note that $f(1, 1) = e$, $f(2, 2) = e^8$, and $f(4, 4) = e^{64}$. If f is homogeneous, it must be the case that $f(2, 2) = 2^t f(1, 1)$ and $f(4, 4) = 2^t f(2, 2)$ for some value t. This requires that $e^8 = 2^t e$ and $e^{64} = 2^t e^8$. Taking ratios, this means that $2^t = e^8/e = e^7$ and $2^t = e^{64}/e^8 = e^{56}$, which are clearly incompatible statements. Thus f cannot be a homogeneous function.

S11.3. **(a)** The function $f(x_1, x_2, x_3) = x_1^{a_1} x_2^{a_2} x_3^{a_3}$ is homogeneous of degree $a_1 + a_2 + a_3$. To see this, note that

$$\begin{aligned}
f(sx_1, sx_2, sx_3) &= (sx_1)^{a_1}(sx_2)^{a_2}(sx_3)^{a_3} \\
&= s^{a_1} x_1^{a_1} s^{a_2} x_2^{a_2} s^{a_3} x_3^{a_3} \\
&= s^{a_1} s^{a_2} s^{a_3} (x_1^{a_1} x_2^{a_2} x_3^{a_3}) \\
&= s^{a_1+a_2+a_3} (x_1^{a_1} x_2^{a_2} x_3^{a_3}) \\
&= s^{a_1+a_2+a_3} f(x_1, x_2, x_3)
\end{aligned}$$

Since any function that is homogeneous is also homothetic, it follows that this function is both homogeneous and homothetic.

(b) The function $g(x_1, x_2, x_3) = x_1^{1/2} x_2^{1/4} x_3^{1/4}$ is homogeneous of degree $1/2 + 1/4 + 1/4 = 1$. To see this, note that

$$\begin{aligned}
g(sx_1, sx_2, sx_3) &= (sx_1)^{1/2}(sx_2)^{1/4}(sx_3)^{1/4} \\
&= s^{1/2} x_1^{1/2} s^{1/4} x_2^{1/4} s^{1/4} x_3^{1/4} \\
&= s^{1/2} s^{1/4} s^{1/4} (x_1^{1/2} x_2^{1/4} x_3^{1/4}) \\
&= s^{1/2+1/4+1/4} (x_1^{1/2} x_2^{1/4} x_3^{1/4}) \\
&= s^1 g(x_1, x_2, x_3)
\end{aligned}$$

Since the function $f(x_1, x_2, x_3) = 1 + g(x_1, x_2, x_3)$ is a monotonic transformation of $g(x_1, x_2, x_3)$, where $g(x_1, x_2, x_3)$ is a homogeneous function, it follows that f is a homothetic function. However, f is not homogeneous. To see this, pick the points $(1, 1, 1)$, $(2, 2, 2)$, and $(4, 4, 4)$. If f is homogeneous of degree k, then it follows that

$$f(2, 2, 2) = 2^k f(1, 1, 1)$$

$$f(4, 4, 4) = 2^k f(2, 2, 2)$$

To see that this is not possible, note that

$$\begin{aligned}
f(1, 1, 1) &= 1 + 1^{1/2}1^{1/4}1^{1/4} = 2 \\
f(2, 2, 2) &= 1 + 2^{1/2}2^{1/4}2^{1/4} \\
&= 1 + 2^{1/2+1/4+1/4} = 1 + 2^1 = 1 + 2 = 3 \\
f(4, 4, 4) &= 1 + 4^{1/2}4^{1/4}4^{1/4} \\
&= 1 + 4^{1/2+1/4+1/4} = 1 + 4^1 = 1 + 4 = 5
\end{aligned}$$

Substituting these results into the conditions above, we get

$$f(2, 2, 2) = 2^k f(1, 1, 1) \Rightarrow 3 = 2^k \times 2$$

and

$$f(4, 4, 4) = 2^k f(2, 2, 2) \Rightarrow 5 = 2^k \times 3$$

The first result implies that $2^k = 3/2$ while the second implies $2^k = 5/3$. This is a contradiction, and so the function f is not homogeneous.

(c) The function $f(x_1, x_2) = x_1^{1/2}x_2^{1/3} + x_2^{3/2}$ is neither homogeneous nor homothetic. To see that it is not homogeneous, we simply note that upon scaling up all x_i values by the factor s, we obtain

$$f(sx_1, sx_2) = s^{5/6}x_1^{1/2}x_2^{1/3} + s^{3/2}x_2^{3/2}$$

which cannot be written as $s^k f(x_1, x_2)$. However, it is not so obvious how to show that f cannot be written as a monotonic transformation of some function that is homogeneous, and so this approach is not so helpful in showing that f is not homothetic. Instead, note that if f is a homothetic function, then it must satisfy the property that the slope of each level curve is equal along any ray from the origin. Thus f_1/f_2 must be equal at points $(k\bar{x}_1, k\bar{x}_2)$ and $(\bar{x}_1, \bar{x}_2)$. To see that this doesn't hold for the function $f(x_1, x_2) = x_1^{1/2}x_2^{1/3} + x_2^{3/2}$, compute the first-order partial derivatives:

$$f_1(x_1, x_2) = \frac{1}{2}x_1^{-1/2}x_2^{1/3} = \frac{\frac{1}{2}x_2^{1/3}}{x_1^{1/2}}$$

$$\begin{aligned} f_2(x_1, x_2) &= \frac{1}{3}x_1^{1/2}x_2^{-2/3} + \frac{3}{2}x_2^{1/2} \\ &= \frac{\frac{1}{3}x_1^{1/2}}{x_2^{2/3}} + \frac{3}{2}x_2^{1/2} \end{aligned}$$

If the function f is homothetic, it follows that the ratio f_1/f_2 must be equal, for example, at the points (1, 1) and (2, 2). To see that this is not so, note that

$$f_1(1, 1) = 1/2 \quad \text{and} \quad f_1(2, 2) = 1/2^{7/6} \doteq 0.445$$

$$f_2(1, 1) = 11/6 \quad \text{and} \quad f_2(2, 2) = \frac{1}{3 \times 2^{1/6}} + \frac{3}{2^{1/2}} \doteq 2.42$$

Thus at the point (1, 1) we get $f_1/f_2 = (1/2)/(11/6) \doteq 0.27$, while at the point (2, 2) we get $f_1/f_2 \doteq 0.445/2.42 \doteq 0.18$. Since these two values are not equal, we know that f is not homothetic.

Chapter S12 Optimization of Functions of *n*-Variables

Contents

S12.1 Price-Discriminating Monopoly with Linear Demands and Costs

A monopoly seller is able to divide its overall market into two submarkets, with the (inverse-) demand functions

$$\begin{aligned} p_1 &= 100 - q_1 \\ p_2 &= 120 - 2q_2 \end{aligned}$$

It produces output at a single plant with the cost function

$$C = 20(q_1 + q_2)$$

Thus unit cost is constant at \$20 per unit, there are no fixed costs, and outputs to the two markets are indistinguishable in production. The firm's profit function is

$$\begin{aligned} \pi(q_1, q_2) &= p_1 q_1 + p_2 q_2 - C \\ &= 100q_1 - q_1^2 + 120q_2 - 2q_2^2 - 20(q_1 + q_2) \\ &= 80q_1 - q_1^2 + 100q_2 - 2q^2 \end{aligned}$$

The firm wishes to choose outputs and prices in the two markets to maximize profit, and so applying theorem 12.1 gives

$$\begin{aligned} \pi_1(q_1^*, q_2^*) &= 80 - 2q_1^* = 0 \\ \pi_2(q_1^*, q_2^*) &= 100 - 4q_2^* = 0 \end{aligned}$$

resulting in outputs and prices of

$$q_1^* = 40, \quad p_1^* = \$60$$
$$q_2^* = 25, \quad p_2^* = \$70$$

and a total profit of $\pi^* = \$2{,}850$. It can be shown that this, in fact, is a true maximum, and we will pursue that result further in the next section.

We now consider the question of the allocation of total output for the price discriminating monopolist. To do this, it is useful to set the problem up a little more generally. Let $R_1(q_1)$ and $R_2(q_2)$ denote the revenue functions in the two markets, and let the cost function be $C = c(q_1 + q_2)$. Then the firm's profit is

$$\pi(q_1, q_2) = R_1(q_1) + R_2(q_2) - c(q_1 + q_2)$$

Applying theorem 12.1, we have the conditions

$$\pi_1(q_1^*, q_2^*) = R_1'(q_1^*) - c = 0$$
$$\pi_2(q_1^*, q_2^*) = R_2'(q_2^*) - c = 0$$

It is then clear that the conditions imply that marginal revenues in the two submarkets must be equalized; that is to say, we must have

$$R_1'(q_1^*) = R_2'(q_2^*)$$

The intuition underlying this condition is easy to see. If marginal revenues in the two markets were unequal, a unit of output could be switched from the market with the lower marginal revenue, to the market with the higher marginal revenue, giving a net increase in revenue, with no increase in cost because total output has remained constant and so profit would increase. This simple condition has some further interesting implications. It is easily shown (see section 6.1) that marginal revenues can always be written as

$$R_i'(q_i) = p_i\left(1 - \frac{1}{\epsilon_i}\right), \qquad i = 1, 2$$

where $\epsilon_i \equiv -(dq_i/dp_i)(p_i/q_i)$ is price elasticity of demand for good i. Then since at the optimal outputs q_i^* marginal revenues are equal, we obtain

$$\frac{p_1}{p_2} = \frac{1 - (1/\epsilon_2)}{1 - (1/\epsilon_1)}$$

From this it can easily be shown that the market with the higher equilibrium price is the one with the lower price-elasticity of demand at the optimal point. We can confirm this in the present example by noting that

$$\epsilon_1 = -(-1)\left(\frac{60}{40}\right) = 1.5, \quad \epsilon_2 = -(-0.5)\left(\frac{70}{25}\right) = 1.4$$

Thus market 2, with the lower demand elasticity at the optimum, has the higher price. As compared to the case in which the same price is set in the two markets, profit is increased by raising price in the market with less elastic demand and lowering it in the market with more elastic demand, until marginal revenues are equalized.

Example S12.1 Cournot Equilibrium with n Identical Firms

Consider an oligopoly with n identical firms, where the market-demand function is now

$$p = 100 - \sum q_i$$

and each firm's profit is

$$\pi_i = 100q_i - q_i \sum q_j - q_i^2, \qquad i, j = 1, \ldots, n;\ i \neq j$$

Find the Cournot equilibrium outputs.

Solution

Maximizing each firm's profit, taking all other outputs as fixed, gives

$$\frac{\partial \pi_i}{\partial q_i} = 100 - 2q_i - \sum q_j = 0$$

so that each firm's reaction function is

$$q_i = \frac{100 - \sum q_j}{2}$$

We now make the plausible assumption, which strictly speaking should be proved, that in the equilibrium all firms' outputs will be the same; that is, as in the previous example, we have a symmetric equilibrium. Denoting this common value by q^* we then have

$$q^* = \frac{100 - (n-1)q^*}{2}$$

which, solving for q^*, gives

$$q^* = \frac{100}{n+1}$$

We can easily confirm our answer in the previous example for $n = 2$. More generally, we can consider the market equilibrium first for $n = 1$. We see that we obtain the monopoly output $q^* = 50$. Then consider the equilibrium output as the number of firms increases. Obviously the output of each individual firm gets smaller

$$\lim_{n\to\infty} q^* = \lim_{n\to\infty}\left(\frac{100}{n+1}\right) = 0$$

but consider the total output $Q^* = nq^*$. We have

$$Q^* = \frac{100n}{n+1} = 100\left(\frac{1}{1+1/n}\right)$$

and so

$$\lim_{n\to\infty} Q^* = 100$$

Thus, as the number of firms increases, total output in the market tends toward the perfectly competitive level (at which p = marginal cost = 0).

The fact that we obtain both monopoly and competitive market outcomes as limiting cases of the Cournot model, as we set $n = 1$ and then let n go to infinity respectively, has undoubtedly contributed to the model's appeal. ■

Example S12.2 Two-Plant Monopoly

A monopoly supplies its markets from two plants, with cost functions

$$C_1 = q_1^2, \quad C_2 = 2q_2$$

and faces a linear market-demand curve

$$p = 70 - 2(q_1 + q_2)$$

Find the firm's profit-maximizing output for each plant.

Solution

The firm's profit is

$$\pi(q_1, q_2) = 70(q_1 + q_2) - 2(q_1 + q_2)^2 - q_1^2 - 2q_2$$

Applying theorem 12.1, we have

$$\frac{\partial \pi}{\partial q_1} = 70 - 6q_1 - 4q_2 = 0$$
$$\frac{\partial \pi}{\partial q_2} = 68 - 4q_1 - 4q_2 = 0$$

which solve to give $q_1 = 1$ and $q_2 = 16$. ■

Example S12.3 **Optimal Input Quantities for a Competitive Firm**

Solve the competitive firm's profit-maximizing use of labor and capital for the case where $y = L^{0.2}K^{0.6}$, $p = 100$, $w = 10$, and $r = 20$. Show that the solution is a true maximum.

Solution

The firm's profit is

$$\pi(L, K) = 100L^{0.2}K^{0.6} - 10L - 20K$$

The first-order conditions are

$$\frac{\partial \pi}{\partial L} = 20L^{-0.8}K^{0.6} - 10 = 0$$
$$\frac{\partial \pi}{\partial K} = 60L^{0.2}K^{-0.4} - 20 = 0$$

Before solving these, we check for a maximum. We have

$$|H_1| = -16\frac{K^{0.6}}{L^{1.8}} < 0 \qquad \text{for any } (K, L) \in \mathbb{R}^2_{++}$$

and

$$|H| = 20(60)L^{-1.6}K^{-0.8}(1 - 0.2 - 0.6) = \frac{240}{L^{1.6}K^{0.8}} > 0 \qquad \text{for any } (K, L) \in \mathbb{R}^2_{++}$$

and so we have a true maximum. Upon dividing the two first-order conditions, we get $L/K = 2/3$ or $L = (2/3)K$ and substitution for L into either first-order condition gives us, after a little algebra, $L^* = 108$ and $K^* = 162$ as the profit-maximizing quantities of inputs. ■

Example S12.4 **Multiproduct Monopoly Revisited**

In the two-output monopoly example in section 12.1, we considered the maximum of the profit function

$$\pi(x_1, x_2) = 78.57x_1 + 57.14x_2 - 0.71x_1 - 0.43x_2 - 0.43x_1x_2 - 50$$

We find that the Hessian matrix of second-order partials is

$$\begin{bmatrix} -1.42 & -0.43 \\ -0.43 & -0.86 \end{bmatrix}$$

It is then straightforward to confirm that the second-order conditions for a maximum are satisfied, since the principal minors are

$$-1.42 < 0; (-1.42)(-0.86) - (-0.43)^2 = 1.0363 > 0$$

Note also that the second-order partials are not functions of x, and so the second-order conditions are satisfied at all values of x (though not, of course, the first-order conditions). This is because, as figure 12.3 showed, the profit function is a strictly concave function. ■

Example S12.5 **Multiplant Monopoly with Linear Costs**

Suppose that a monopoly supplies its market from two plants, with cost functions:

$$C_1 = 5q_1, \quad C_2 = 6q_2 \tag{S12.1}$$

This means that a unit of output costs \$5 to produce at plant 1 and \$6 at plant 2, and unit cost does not vary with output. Given the linear demand

$$p = 100 - (q_1 + q_2) \tag{S12.2}$$

the firm's profit function is

$$\pi = 100(q_1 + q_2) - (q_1 + q_2)^2 - 5q_1 - 6q_2$$

We wish to find the profit-maximizing output from each plant. Applying theorem 12.1, we have

$$\frac{\partial \pi}{\partial q_1} = 100 - 2q_1 - 2q_2 - 5 = 0$$
$$\frac{\partial \pi}{\partial q_2} = 100 - 2q_2 - 2q_1 - 6 = 0$$

These conditions then give the equations

$$2q_1 + 2q_2 = 95, \quad 2q_1 + 2q_2 = 94$$

which, of course, have no solution. The lines defined by the equations are parallel. What went wrong?

The answer is easy to see. Plant 2's unit cost, at \$6, is always greater than plant 1's unit cost, at \$5. So, it would never pay to use plant 2; we should simply set its output at zero and find the profit-maximizing output at plant 1. From the preceding first-order condition with $q_2 = 0$, this gives $q_1 = 95/2 = 47.5$. It turns out that introducing nonnegativity conditions explicitly, meaning that $q_i \geq 0$, $i = 1, 2$, resolves the problem. These of course are perfectly reasonable restrictions to impose in any case, but in this problem they are crucial. Thus we reformulate the problem as

$$\max \pi(q_1, q_2) = 100(q_1 + q_2) - (q_1 + q_2)^2 - 5q_1 - 6q_2 \quad \text{s.t.} \quad 0 \leq q_1, 0 \leq q_2$$

This is clearly a special case of the problem considered in theorem 12.7, with $a_i = 0$ and b_i at $+\infty$. We therefore need apply only (i) of the theorem. The profit-maximizing outputs q_i must satisfy

$$\pi_1(q_1^*, q_2^*) = 100 - 2(q_1^* + q_2^*) - 5 \leq 0 \text{ and } q_1^*(100 - 2(q_1^* + q_2^*) - 5) = 0$$
$$\pi_2(q_1^*, q_2^*) = 100 - 2(q_1^* + q_2^*) - 6 \leq 0 \text{ and } q_2^*(100 - 2(q_1^* + q_2^*) - 6) = 0$$

There are three solution possibilities (excluding the case where both outputs are zero):

(a) $q_1 > 0, q_2 > 0$
(b) $q_1 > 0, q_2 = 0$
(c) $q_1 = 0, q_2 > 0$

We now show that only case (b) is possible. Suppose that we have case (a). Then from the conditions we must have

$$\begin{aligned} 100 - 2(q_1 + q_2) &= 5 \\ 100 - 2(q_1 + q_2) &= 6 \end{aligned}$$

This is of course impossible, and so we cannot have case (a). Suppose that we now have case (c). Then the conditions imply that

$$\begin{aligned} 100 - 2(q_1 + q_2) &\leq 5 \\ 100 - 2(q_1 + q_2) &= 6 \end{aligned}$$

which again is impossible. Suppose finally that we have case (b). Then we have

$$\begin{aligned} 100 - 2(q_1 + q_2) &= 5 \\ 100 - 2(q_1 + q_2) &\leq 6 \end{aligned}$$

which give no problem as long as we take the inequality $<$ in the second condition. Therefore we have the solution $q_1 = 47.5$, $q_2 = 0$.

Of course, the answer to the problem is always obvious, q_2 should clearly be set to zero. However, methods that seem excessively long-winded in easy problems are very powerful in helping us solve harder ones. We hope to show this in another economic application below. First, however, we consider an extension of the present example. ■

Example S12.6 **Two-Plant Monopoly with Capacity Constraints**

Suppose that we have a two-plant monopoly with the linear cost and demand functions given in equations (S12.1) and (S12.2), and the added feature that each plant has a maximum capacity of 30 units of output. Solve the profit-maximization problem.

Solution

The solution we found previously to this problem, of $q_1 = 47.5$ is no longer feasible because it exceeds the available capacity of plant 1. We must now formulate the firm's profit-maximization problem as

$$\begin{aligned} \max \pi(q_1, q_2) &= 100(q_1 + q_2) - (q_1 + q_2)^2 - 5q_1 - 6q_2 - 50 \\ \text{s.t.} \quad & 0 \leq q_1 \leq 30, 0 \leq q_2 \leq 30 \end{aligned}$$

Applying theorem 12.7, we have the conditions

$$\begin{aligned}
\pi_1(q_1^*, q_2^*) &= 100 - 2(q_1^* + q_2^*) - 5 \leq 0 \\
\text{and} \quad q_1^*(100 - 2(q_1^* + q_2^*) - 5) &= 0 \\
\pi_1(q_1^*, q_2^*) &= 100 - 2(q_1^* + q_2^*) - 5 \geq 0 \\
\text{and} \quad (30 - q_1^*)(100 - 2(q_1^* + q_2^*) - 5) &= 0
\end{aligned}$$

$$\begin{aligned}
\pi_2(q_1^*, q_2^*) &= 100 - 2(q_1^* + q_2^*) - 6 \leq 0 \\
\text{and} \quad q_2^*(100 - 2(q_1^* + q_2^*) - 6) &= 0 \\
\pi_2(q_1^*, q_2^*) &= 100 - 2(q_1^* + q_2^*) - 6 \geq 0 \\
\text{and} \quad (30 - q_2^*)(100 - 2(q_1^* + q_2^*) - 6) &= 0
\end{aligned}$$

We have in total nine logical possibilities, since each output may be within, or at one of the end points of its interval. It will save a lot of tedium, however, if we use what we already know to rule some of these out from the start. We know that q_1 will be at its capacity level, which gives us the condition

$$100 - 2(30 + q_2^*) - 5 \geq 0$$

We can rule out immediately the case in which $q_2^* = 30$, because from the last condition this would give us

$$100 - 2(30 + 30) - 6 \geq 0$$

which is clearly false. We can also rule out the case in which $q_2^* = 0$ because this would give us the condition

$$100 - 2(30 + 0) - 6 \leq 0$$

which is also false. So we must have $30 > q_2^* > 0$, in which case we have the condition

$$100 - 2(30 + q_2^*) - 6 = 0$$

giving the solution $q_2^* = 17$. As a check, note that

$$100 - 2(30 + 17) - 5 = 1 \geq 0$$

and so there is no contradiction with the condition on q_1^*.

This solution may seem needlessly long-winded. If plant 1 is being used to capacity, then the firm's marginal cost is \$6, which is the marginal cost of plant 2.

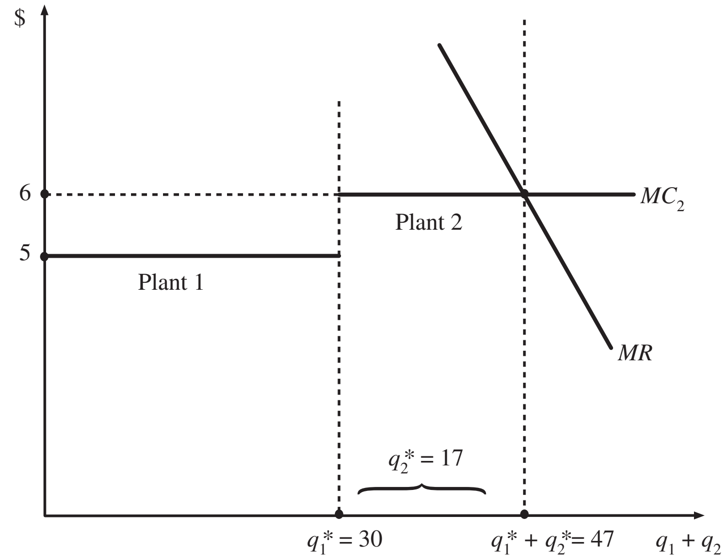

Figure S12.1 Two-plant monopoly with fixed capacities

Then we could simply have solved the problem

$$\max 100q - q^2 - 6q - 50$$

where q is total output. We then obtain the condition

$$94 - 2q^* = 0$$

giving the solution $q^* = 47$, of which we know 30 will be produced in plant 1. Figure S12.1 illustrates the simple solution.

However, although in simple problems there may be more direct ways to the solution than the grinding through of logical possibilities given by the conditions in theorem 12.5, in more complex cases these turn out to be very valuable. ■

Chapter S13 Constrained Optimization

Contents

S13.1 A Farmer's Land Allocation

A farmer has a given amount of land, denoted by $\bar{l}$, and can allocate it between two crops. l_i, $i = 1, 2$, is the amount of land allocated to crop i. Each crop is sold on a competitive market at a given price. The production functions for the crops are given by

$$y_i = l_i^{a_i}, \qquad a_i \in (0, 1),\ i = 1, 2$$

where y_i is output of crop i. For simplicity, we assume that the required amounts of labor, fertilizer, and so on, are fixed quantities per unit of land and so need not be chosen separately. The net profit per unit of output, price minus variable costs (assumed also constant per unit of output), is denoted by r_i, $i = 1, 2$. Next we assume that the land has no alternative use or market: it will be used for growing these two crops or not at all. If the farmer wants to maximize his net income, he must solve

$$\max\ r_1 y_1 + r_2 y_2 \quad \text{s.t.} \quad \bar{l} - l_1 - l_2 = 0 \quad \text{and} \quad y_i = l_i^{a_i}, \quad i = 1, 2$$

We can proceed in one of three ways with this problem. We could treat it as a problem with four variables and three constraints, but we have not yet made the extension to the theory required to handle this case. Or, we could substitute for the outputs in the profit function to obtain

$$\max\ r_1 l_1^{a_1} + r_2 l_2^{a_2} \quad \text{s.t.} \quad \bar{l} - l_1 - l_2 = 0$$

which is a problem in two land use variables. Or, we could invert the production functions to obtain the *input requirement functions*

$$l_i = y_i^{1/a_i}, \qquad i = 1, 2$$

and substitute into the constraint to obtain

$$\max\ r_1 y_1 + r_2 y_2 \quad \text{s.t.} \quad \bar{l} - y_1^{1/a_1} - y_2^{1/a_2} = 0$$

Each method would lead to the same solution, but each gives a somewhat different insight into the nature of the problem. Here we will work with the l_i and leave the last formulation as an exercise. Thus the Lagrange function is

$$\mathcal{L} = r_1 l_1^{a_1} + r_2 l_2^{a_2} + \lambda(\bar{l} - l_1 - l_2)$$

and the first-order conditions are

$$\frac{\partial \mathcal{L}}{\partial l_1} = r_1 a_1 l_1^{a_1 - 1} - \lambda = 0$$
$$\frac{\partial \mathcal{L}}{\partial l_2} = r_2 a_2 l_2^{a_2 - 1} - \lambda = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = \bar{l} - l_1 - l_2 = 0$$

We can interpret $r_i a_i l_i^{a_i - 1}$ as the *marginal net profit of land* allocated to crop i. It is the product of marginal net profit per unit of output, r_i, and the marginal physical product of land producing crop i, $a_i l_i^{a_i - 1}$. Since $a_i < 1$, this marginal physical product is positive but decreasing in l_i.

Then, using the first two conditions, we immediately have that the optimal allocation of land is characterized by the condition

$$r_1 a_1 l_1^{a_1 - 1} = r_2 a_2 l_2^{a_2 - 1}$$

The profit earned from the marginal bit of land allocated to each crop must be equal, so that total profit could not be increased by reallocating a little land from one crop to the other. Figure S13.1 illustrates. The distance $0L$ on the horizontal axis gives the total amount of land available, l_1 is measured rightward from 0, and l_2 is measured leftward from L.

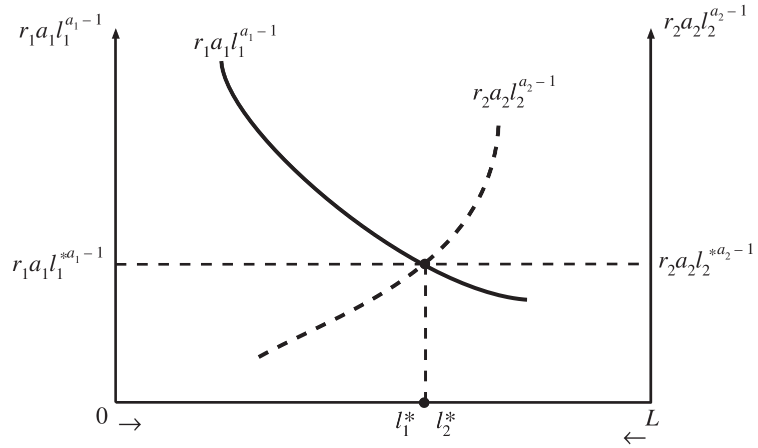

Figure S13.1 Optimal land allocation

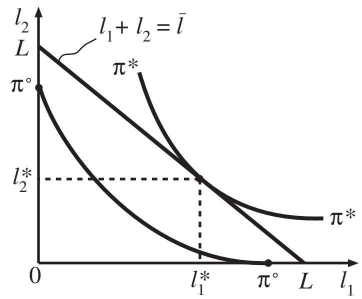

Figure S13.2 Optimal land allocation as a tangency solution

To interpret the solution in terms of the tangency of the level curves of the constraint and objective functions, first note that the condition above can be rearranged to give

$$\frac{r_1a_1l_1^{a_1-1}}{r_2a_2l_2^{a_2-1}} = 1$$

In figure S13.2 we show the constraint as the line LL, and its slope is -1. We define the level curves of the objective function by setting

$$r_1l_1^{a_1} + r_2l_2^{a_2} = \pi$$

For a given π the typical level curve is shown by $\pi_0\pi_0$. The slope of the level curve is

$$-\frac{r_1a_1l_1^{a_1-1}}{r_2a_2l_2^{a_2-1}}$$

Thus we have a tangency solution at l_1^* and l_2^*.

Example S13.1 **Numerical Version of the Land-Allocation Problem**

The farmer's available amount of land is 1,000 acres, the unit net profit of crops 1 and 2 are \$10 and \$8 respectively, and the production functions are

$$y_1 = l_1^{0.6}, \quad y_2 = l_2^{0.8}$$

What is the optimal land allocation?

Solution

The Lagrange function is

$$\mathcal{L} = 10l_1^{0.6} + 8l_2^{0.8} + \lambda(1{,}000 - l_1 - l_2)$$

and the first-order conditions are

$$\begin{aligned} 6l_1^{-0.4} - \lambda &= 0 \\ 6.4l_2^{-0.2} - \lambda &= 0 \\ 1{,}000 - l_1 - l_2 &= 0 \end{aligned}$$

The first two conditions give

$$6l_1^{-0.4} = 6.4l_2^{-0.2}$$

implying that

$$6l_2^{0.2} = 6.4l_1^{0.4}$$

and so

$$l_2 = \left[\left(\frac{6.4}{6}\right)l_1^{0.4}\right]^5 = 1.38l_1^2$$

Substituting into the constraint gives the quadratic equation

$$1.38l_1^2 + l_1 - 1{,}000 = 0$$

This solves to give

$$l_1^* = 26.6, \quad l_2^* = 973.4$$

Note how much more important is the greater marginal productivity in producing y_2 than the higher net profit per unit of output of y_1. ■

S13.2 Consumer-Demand Functions with Cobb-Douglas Utility

The standard model of the consumer in economic theory is concerned with deriving the consumer's demands for goods as a constrained maximization problem. In the two-good case, the consumer has a utility function $u(x_1, x_2)$, defined on bundles of goods (x_1, x_2), and she is assumed to maximize this utility, subject to the constraint that the amount she spends on the goods, given by the sum of prices times quantities, $p_1x_1 + p_2x_2$, cannot exceed the income she has available, m. Therefore we write the consumer's problem as

$$\max\ u(x_1, x_2) \quad \text{s.t.} \quad m - p_1x_1 - p_2x_2 = 0$$

We assume that the consumer's preferences are such that the utility function takes the Cobb-Douglas form

$$u = x_1^{\alpha}x_2^{1-\alpha} \qquad \text{with } 0 < \alpha < 1$$

We want to examine for this utility function the form of the demand functions that shows how the consumer's demands depend on prices and income. We write the Lagrange function

$$\mathcal{L}(x_1, x_2, \lambda) = x_1^{\alpha}x_2^{1-\alpha} + \lambda(m - p_1x_1 - p_2x_2)$$

Then we apply definition 13.1 to obtain

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial x_1} &= \alpha x_1^{\alpha-1}x_2^{1-\alpha} - \lambda p_1 = 0 \\
\frac{\partial \mathcal{L}}{\partial x_2} &= (1-\alpha)x_1^{\alpha}x_2^{-\alpha} - \lambda p_2 = 0 \\
\frac{\partial \mathcal{L}}{\partial \lambda} &= m - p_1x_1 - p_2x_2 = 0
\end{aligned}$$

We wish to derive solutions for x_1 and x_2 as functions of p_1, p_2, and m. To do this, we first eliminate λ from the first two conditions, to obtain

$$\frac{\alpha x_1^{\alpha-1}x_2^{1-\alpha}}{(1-\alpha)x_1^{\alpha}x_2^{-\alpha}} = \frac{p_1}{p_2}$$

Simplifying gives

$$\frac{\alpha x_2}{(1-\alpha)x_1} = \frac{p_1}{p_2}$$

This is the condition that the slope of the indifference curve be equal to the slope of the budget constraint. We can solve for, say, x_2 as a function of x_1 to obtain

$$x_2 = \left[\frac{(1-\alpha)p_1}{\alpha p_2}\right]x_1$$

Then substituting into the budget constraint gives

$$m - p_1x_1 - p_2\left[\frac{(1-\alpha)p_1x_1}{\alpha p_2}\right] = m - p_1\left[1 + \frac{1-\alpha}{\alpha}\right]x_1 = 0$$

Solving for x_1 gives the demand function for x_1:

$$x_1 = \frac{\alpha m}{p_1} \tag{S13.1}$$

Substituting for x_1 in the expression for x_2 gives the demand function for x_2:

$$x_2 = \frac{(1-\alpha)m}{p_2} \tag{S13.2}$$

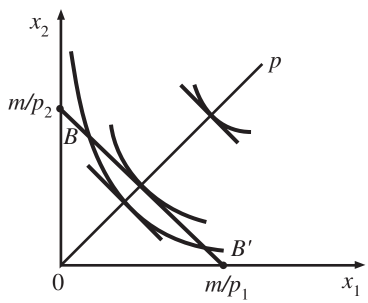

Figure S13.3 Solution to the consumer's demand problem

In figure S13.3 we graph the level curves of the utility function, the consumer's indifference curves. The line BB' graphs the budget constraint, and so has slope $-p_1/p_2$ and intercepts m/p_1 and m/p_2. Points on this line require an expenditure exactly equal to the consumer's income. The solution is a point of tangency, and the expressions given above for x_1 and x_2 enable us to calculate these solution values once we have numerical values for the parameters α, p_1, and p_2.

The demand functions (13.4) and (13.5) have a number of interesting properties:

1. The demand curves drawn from them are negatively sloped with a constant own-price elasticity of 1. Thus differentiating the first demand function gives

$$\frac{\partial x_1}{\partial p_1} = \frac{-\alpha m}{p_1^2}$$

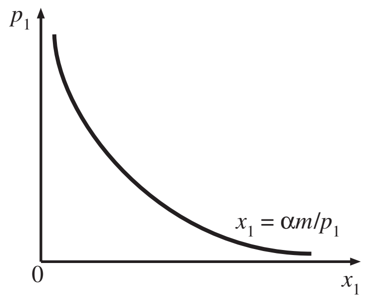

Figure S13.4 The demand curve for a Cobb-Douglas utility function is a rectangular hyperbola

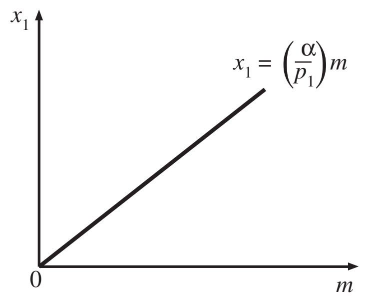

Figure S13.5 The Engel curve for a Cobb-Douglas utility function is linear

and so

$$e_1 = -\left(\frac{p_1}{x_1}\right)\frac{\partial x_1}{\partial p_1} = -\left(\frac{p_1^2}{\alpha m}\right)\left(\frac{-\alpha m}{p_1^2}\right) = 1$$

and similarly for the second demand function. In fact, when graphed, these functions are *rectangular hyperbolas* (see figure S13.4).

2. Multiply the first demand function through by p_1. We then have

$$p_1 x_1 = \alpha m$$

Thus at every price, demand is such that precisely the same proportion of income is spent on the good. For good 1, this proportion is α, while for good 2 the proportion is $1 - \alpha$.

3. Good 1's **Engel curve,** which relates quantity demanded to income, is a straight line through the origin with slope α/p_1 (see figure S13.5). The elasticity of expenditure with respect to income (denoted χ_1) is 1, as is the income elasticity of demand (η_1):

$$\chi_1 = \left(\frac{\partial(p_1 x_1)}{\partial m}\right)\left(\frac{m}{p_1 x_1}\right) = \alpha\left(\frac{1}{\alpha}\right) = 1$$

$$\eta_1 = \left(\frac{\partial x_1}{\partial m}\right)\left(\frac{m}{x_1}\right) = \left(\frac{\alpha}{p_1}\right)\left(\frac{p_1}{\alpha}\right) = 1$$

Similarly for good 2.

4. Each good's demand is independent of the other good's price. This is easy to see, since p_1 does not appear in the demand function for x_2, and vice versa.

This example shows how, given a specific utility function, we can solve the consumer's constrained maximization problem to obtain demand functions with specific properties. Other functions are of course possible, though few are as easy to work with!

Example S13.2 In equations (S13.1) and (S13.2) if $\alpha = 0.25$, and $m = 100$, then for prices p_1 and p_2 the demand curves are

$$x_1 = \frac{25}{p_1}, \quad x_2 = \frac{75}{p_2}$$

which have properties 1 to 4 above. ■

S13.3 Long-Run Cost Function for a Firm with Cobb-Douglas Production Function

If a firm wishes to maximize its profit, then the implication is that it will want to minimize the cost of producing any given level of output. In fact the standard model of the profit-maximizing firm proceeds in two steps: we first find the firm's cost function, expressing *minimized cost* as a function of the output level and input prices; then we find the profit-maximizing level of output by combining the firm's cost and revenue functions. Here we are concerned with the first of these stages, using for concreteness a specific functional form for the production function.

Assume that the firm uses two inputs, labor L, and capital K. The firm's production possibilities are described by the Cobb-Douglas production function

$$y = K^{\alpha}L^{\beta}, \qquad \alpha, \beta > 0$$

where y is output. If r is the price of a unit of capital and w the price of a unit of labor, then the firm's total costs are

$$C = rK + wL$$

If the firm wishes to minimize the cost of producing some given level of output y, then it has to solve the constrained minimization problem

$$\min\ rK + wL \quad \text{s.t.} \quad y - K^{\alpha}L^{\beta} = 0$$

The Lagrange function for the problem is

$$\mathcal{L}(K, L, \lambda) = rK + wL + \lambda(y - K^{\alpha}L^{\beta})$$

and the first-order conditions are

$$\frac{\partial \mathcal{L}}{\partial K} = r - \lambda\alpha K^{\alpha-1}L^{\beta} = 0$$
$$\frac{\partial \mathcal{L}}{\partial L} = w - \lambda\beta K^{\alpha}L^{\beta-1} = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = y - K^{\alpha}L^{\beta} = 0$$

Eliminating λ from the first two conditions gives

$$\frac{\alpha K^{\alpha-1}L^{\beta}}{\beta K^{\alpha}L^{\beta-1}} = \frac{r}{w}$$

The term on the left-hand side is the marginal rate of technical substitution (MRTS) between the inputs (see example S11.22), and the condition says that cost minimization requires that the MRTS be equated to the input price ratio. The MRTS can be simplified to obtain

$$\frac{\alpha L}{\beta K} = \frac{r}{w}$$

Solving for, say, L as a function of K gives

$$L = \left(\frac{\beta r}{\alpha w}\right) K$$

and then substituting into the constraint gives

$$y - K^{\alpha}\left(\frac{\beta r K}{\alpha w}\right)^{\beta-\alpha} = y - K^{\alpha+\beta}\left(\frac{r}{w}\right)^{\beta-\alpha}\left(\frac{\beta}{\alpha}\right)^{\beta-\alpha} = 0$$

We can then solve for K:

$$K = \left(\frac{\alpha}{\beta}\right)^{\beta/(\alpha+\beta)}\left(\frac{w}{r}\right)^{\beta/(\alpha+\beta)} y^{1/(\alpha+\beta)}$$

This is a demand function for K, since it shows how the firm's desired capital input varies with input prices and planned output level. Similarly

$$L = \left(\frac{\beta}{\alpha}\right)^{\alpha/(\alpha+\beta)}\left(\frac{r}{w}\right)^{\alpha/(\alpha+\beta)} y^{1/(\alpha+\beta)}$$

is the demand function for labor.

We illustrate this solution in figure S13.6. The curve labeled y is a level curve of the production function corresponding to the given output level, which is called an *isoquant* of the production function (see section 11.3). It is the constraint curve in this problem, since we are required to choose from input pairs that produce the required output level. At any given cost level C, the input pairs that incur that level of cost lie along a straight line with slope $-w/r$, and intercepts C/w and C/r. The higher is C, the higher is the cost line, and so the problem of *minimizing* cost is, diagrammatically speaking, the problem of getting onto the *lowest* possible cost line. The solution is then at the point of tangency shown.

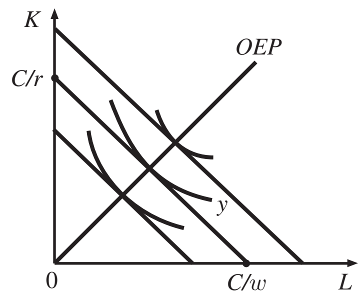

Figure S13.6 Solution to the firm's cost-minimization problem

Consider now the straight line labeled *OEP* in figure S13.6. Since it passes through the tangency point, it has the equation

$$K = \left(\frac{\alpha w}{\beta r}\right) L$$

and shows the set of pairs of (L, K)-values that satisfy the tangency condition at given prices as y varies. That is, if we were to change the required output level and re-solve the constrained minimization problem with the same prices, we would always obtain a solution on OEP. This line is called the **output expansion path,** because it shows how the *cost-minimizing* input pairs change as required output expands. It is a particular feature of this example that the output expansion path is a straight line through the origin.

Returning to the algebra, we can derive the cost function in the following way. Given the cost equation $C = rK + wL$, we obtain *minimized* cost as a function of output and input prices by substituting in the optimal solutions for K and L. Therefore, at the optimal input combination for the given output level, total cost will be

$$\begin{aligned} C &= rK + wL \\ &= r\left(\frac{\alpha}{\beta}\right)^{\beta/(\alpha+\beta)}\left(\frac{w}{r}\right)^{\beta/(\alpha+\beta)} y^{1/(\alpha+\beta)} \\ &\quad + w\left(\frac{\beta}{\alpha}\right)^{\alpha/(\alpha+\beta)}\left(\frac{r}{w}\right)^{\alpha/(\alpha+\beta)} y^{1/(\alpha+\beta)} \end{aligned}$$

This expression can be simplified readily. Let

$$a = \left(\frac{\alpha}{\beta}\right)^{\beta/(\alpha+\beta)} + \left(\frac{\beta}{\alpha}\right)^{\alpha/(\alpha+\beta)}$$

and note that

$$r\left(\frac{w}{r}\right)^{\beta/(\alpha+\beta)} = r^{1-\beta/(\alpha+\beta)} w^{\beta/(\alpha+\beta)} = r^{\alpha/(\alpha+\beta)} w^{\beta/(\alpha+\beta)}$$

$$w\left(\frac{r}{w}\right)^{\alpha/(\alpha+\beta)} = r^{\alpha/(\alpha+\beta)} w^{1-\alpha/(\alpha+\beta)} = r^{\alpha/(\alpha+\beta)} w^{\beta/(\alpha+\beta)}$$

where we have simply used the fact that $1 = (\alpha + \beta)/(\alpha + \beta)$. We can then rearrange the expression above to obtain the cost function

$$C = ar^{\alpha/(\alpha+\beta)} w^{\beta/(\alpha+\beta)} y^{1/(\alpha+\beta)}$$

We note that at any given output, costs are increasing in the input prices, but since we are interested mainly in the relation between cost and output, we suppress the input prices by setting

$$b = ar^{\alpha/(\alpha+\beta)} w^{\beta/(\alpha+\beta)}$$

and writing the cost function as

$$C = by^{1/(\alpha+\beta)}$$

We now want to examine the properties of this cost function. Clearly, its shape depends on the value of the sum $\alpha + \beta$. We have three possibilities:

Case 1 $\alpha + \beta > 1$. In this case, $1/(\alpha + \beta) < 1$, and so the function is strictly concave. The *marginal*- and *average-cost functions* are

$$\text{MC} = \frac{dC}{dy} = \gamma y^{1/(\alpha+\beta)-1}, \quad \text{AC} = \frac{C}{y} = by^{1/(\alpha+\beta)-1}$$

where $\gamma = b/(\alpha + \beta)$. Here the exponents in both the MC and AC functions are negative, implying that the MC and AC curves are both negatively sloped, with the MC curve below the AC curve, since $\gamma < b$. Figure S13.7 illustrates. In economics, the cost function in this case is said to exhibit *economies of scale*.

Case 2 $\alpha + \beta = 1$. In this case, $1/(\alpha + \beta) = 1$, and so the cost function is simply $C = by$, a straight line through the origin. We have MC = AC = b, as figure S13.8 illustrates.

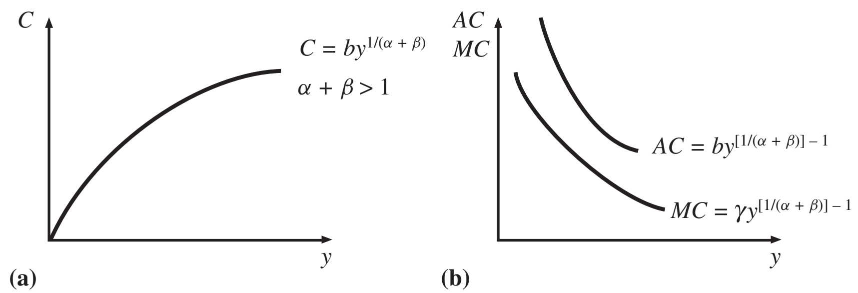

Figure S13.7 Total-, average-, and marginal-cost curves with economies of scale

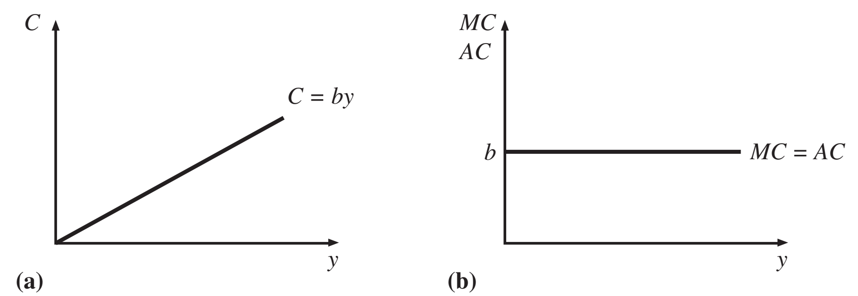

Figure S13.8 Total-, average-, and marginal-cost curves with constant returns to scale

Case 3 $\alpha + \beta < 1$. In this case, $1/(\alpha + \beta) > 1$, and the cost function is strictly convex. The exponents in the AC and MC functions are both positive, and so the AC and MC curves are positively sloped, with MC above AC because now $\gamma > b$. The cost function is said to exhibit *diseconomies of scale*. Figure S13.9 illustrates. (Refer to section 11.5 for a more comprehensive discussion of returns to scale.)

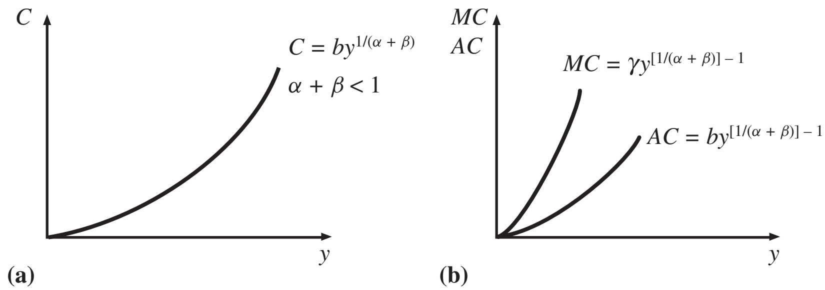

Figure S13.9 Total-, average-, and marginal-cost curves for diseconomies of scale

S13.4 Cost Minimization with the CES Production Function

In chapter 11 we defined the CES production function as

$$y = [\delta L^{-r} + (1-\delta)k^{-r}]^{-1/r}, \qquad 1 > \delta > 0, r > -1$$

where the elasticity of substitution σ is given by

$$\sigma \equiv \frac{1}{1+r}$$

For present purposes, in light of the manipulations necessary to derive the cost function, it is useful to rewrite the CES production function as follows. First, let

$$\rho \equiv -r$$

and then define

$$\alpha_1 = \delta^{1/\rho}, \quad \alpha_2 = (1-\delta)^{1/\rho}$$

We can now write the CES production function in more convenient form

$$y = [(\alpha_1 L)^\rho + (\alpha_2 K)^\rho]^{1/\rho}$$

The problem is

$$\min C = wL + rK \quad \text{s.t.} \quad q = [(\alpha_1 L)^\rho + (\alpha_2 K)^\rho]^{1/\rho}$$

The first-order conditions are

$$w - \lambda \frac{1}{\rho}[(\alpha_1 L)^\rho + (\alpha_2 K)^\rho]^{(1/\rho)-1} \rho(\alpha_1 L)^{\rho-1}\alpha_1 = 0$$

$$r - \lambda \frac{1}{\rho}[(\alpha_1 L)^\rho + (\alpha_2 K)^\rho]^{(1/\rho)-1} \rho(\alpha_2 K)^{\rho-1}\alpha_2 = 0$$

$$y - [(\alpha_1 L)^\rho + (\alpha_2 K)^\rho]^{1/\rho} = 0$$

Eliminating the Lagrange multiplier gives

$$\frac{w}{r} = \left(\frac{\alpha_1}{\alpha_2}\right)^\rho \left(\frac{L}{K}\right)^{\rho-1}$$

Solving for K gives

$$K = \left(\frac{\alpha_1}{\alpha_2}\right)^{\rho/(\rho-1)} \left(\frac{r}{w}\right)^{1/(\rho-1)} L$$

Substituting for K in the production function gives

$$\begin{aligned}
y &= \left[(\alpha_1 L)^\rho + \left(\alpha_2 \left(\frac{\alpha_1}{\alpha_2}\right)^{\rho/(\rho-1)} \left(\frac{r}{w}\right)^{1/(\rho-1)} L\right)^\rho\right]^{1/\rho} \\
&= \left[\left(\alpha_1^\rho + \alpha_2^\rho \alpha_2^{-\rho^2/(\rho-1)} \alpha_1^{\rho^2/(\rho-1)} r^{\rho/(\rho-1)} w^{-\rho/(\rho-1)}\right) L^\rho\right]^{1/\rho} \\
&= \left[\left(\alpha_1^\rho \alpha_1^{-\rho^2/(\rho-1)} w^{\rho/(\rho-1)} + \alpha_2^{\rho-[\rho^2/(\rho-1)]} r^{\rho/(\rho-1)}\right) L^\rho \alpha_1^{\rho^2/(\rho-1)} w^{-\rho/(\rho-1)}\right]^{1/\rho}
\end{aligned}$$

Now note that

$$\rho - \frac{\rho^2}{\rho - 1} = -\frac{\rho}{\rho - 1}$$

and so we have

$$y = \left[\left(\frac{w}{\alpha_1}\right)^{\rho/(\rho-1)} + \left(\frac{r}{\alpha_2}\right)^{\rho/(\rho-1)}\right]^{1/\rho} L\alpha_1^{\rho/(\rho-1)} w^{-1/(\rho-1)}$$

Thus solving for L gives the labor-demand function

$$L = \left[\left(\frac{w}{\alpha_1}\right)^{\rho/(\rho-1)} + \left(\frac{r}{\alpha_2}\right)^{\rho/(\rho-1)}\right]^{-1/\rho} \alpha_1^{-\rho/(\rho-1)} w^{1/(\rho-1)} y$$

In a similar way we find the demand function for capital

$$K = \left[\left(\frac{w}{\alpha_1}\right)^{\rho/(\rho-1)} + \left(\frac{r}{\alpha_2}\right)^{\rho/(\rho-1)}\right]^{-1/\rho} \alpha_2^{-\rho/(\rho-1)} r^{1/(\rho-1)} y$$

To derive the cost function, we must substitute for L and K in the cost equation. First note that

$$\begin{aligned} wL &= \left[\left(\frac{w}{\alpha_1}\right)^{\rho/(\rho-1)} + \left(\frac{r}{\alpha_2}\right)^{\rho/(\rho-1)}\right]^{-1/\rho} \alpha_1^{-\rho/(\rho-1)} w^{\rho/(\rho-1)} y \\ &= \left[\left(\frac{w}{\alpha_1}\right)^{\rho/(\rho-1)} + \left(\frac{r}{\alpha_2}\right)^{\rho/(\rho-1)}\right]^{-1/\rho} \left(\frac{w}{\alpha_1}\right)^{\rho/(\rho-1)} y \end{aligned}$$

Likewise we have

$$rK = \left[\left(\frac{w}{\alpha_1}\right)^{\rho/(\rho-1)} + \left(\frac{r}{\alpha_2}\right)^{\rho/(\rho-1)}\right]^{-1/\rho} \left(\frac{r}{\alpha_2}\right)^{\rho/(\rho-1)} y$$

Then adding these expressions gives

$$\begin{aligned} C &= wL + rK \\ &= \left[\left(\frac{w}{\alpha_1}\right)^{\rho/(\rho-1)} + \left(\frac{r}{\alpha_2}\right)^{\rho/(\rho-1)}\right]^{1-(1/\rho)} y \\ &= \left[\left(\frac{w}{\alpha_1}\right)^{\rho/(\rho-1)} + \left(\frac{r}{\alpha_2}\right)^{\rho/(\rho-1)}\right]^{(\rho-1)/\rho} y \\ &= c(w, r)y \end{aligned}$$

The cost function can be factored into output multiplied by a unit-cost function $c(w, r)$ that depends only on input prices.

Example S13.3 Numerical Version of the CES Cost Function Problem

A firm pays \$10 per unit for input x_1 and \$8 per unit for input x_2. It has the CES production function

$$y = \left(0.4x_1^{-2} + 0.6x_2^{-2}\right)^{-0.5}$$

What is its cost-minimizing input combination to produce one unit of output?

Solution

The Lagrange function is

$$\mathcal{L} = 10x_1 + 8x_2 + \lambda\left[1 - \left(0.4x_1^{-2} + 0.6x_2^{-2}\right)^{-0.5}\right]$$

The first-order conditions are

$$\begin{aligned} 10 - \lambda 0.5\left[0.4x_1^{-2} + 0.6x_2^{-2}\right]^{-1.5} 0.8x_1^{-3} &= 0 \\ 8 - \lambda 0.5\left[0.4x_1^{-2} + 0.6x_2^{-2}\right]^{-1.5} 1.2x_2^{-3} &= 0 \\ 1 - \left[0.4x_1^{-2} + 0.6x_2^{-2}\right]^{-0.5} &= 0 \end{aligned}$$

Taking the ratio of the first two conditions gives

$$\frac{10}{8} = \frac{0.8}{1.2}\left(\frac{x_1}{x_2}\right)^{-3}$$

giving

$$1.875 = \left(\frac{x_2}{x_1}\right)^{3}$$

and so

$$x_2 = (1.875)^{1/3}, \quad x_1 = 1.233x_1$$

Substituting into the constraint gives

$$\left[0.4x_1^{-2} + 0.6(1.233x_1)^{-2}\right]^{-0.5} = 1$$

implying that

$$\left[0.4 + \frac{0.6}{(1.233)^2}\right]x_1^{-2} = 1^{-2} = 1$$

or

$$0.795x_1^{-2} = 1$$

giving, finally,

$$x_1 = \left(\frac{1}{0.795}\right)^{-0.5} = (0.795)^{0.5} = 0.89 \text{ units}$$

We then have

$$x_2^* = 1.1, \quad x_1^* = 0.89 \text{ units}$$

■

S13.5 Optimization with More Than One Constraint

So far in this chapter, we have considered constrained maximization and minimization problems only for the case of one constraint and functions of two variables. This was, however, simply for ease of illustration. The Lagrange procedure extends readily to the case in which there are $m \geq 1$ constraints and n choice variables.

Definition S13.1 The Lagrange method of finding a solution $\mathbf{x}^*$ to the problem

$$\max f(\mathbf{x}) \quad \text{s.t.} \quad \begin{cases} g^1(\mathbf{x}) = 0 \\ g^2(\mathbf{x}) = 0 \\ \dots\dots\dots \\ g^m(\mathbf{x}) = 0 \end{cases}$$

consists of deriving the conditions for a stationary value of the Lagrange function

$$\mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) + \sum_j \lambda_j g^j(\mathbf{x})$$

which are the $n + m$ conditions

$$\frac{\partial \mathcal{L}}{\partial x_i} = f_i(\mathbf{x}^*) + \sum_j \lambda_j^* g_i^j(\mathbf{x}^*) = 0, \qquad i = 1, \ldots, n$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_j} = g^j(\mathbf{x}^*) = 0, \qquad j = 1, \ldots, m$$

One restriction we have to place is that $n > m$. The reason for this is that if $n = m$, then there may be only one feasible point, in which case optimization is unnecessary since we have no choice for the solution. If $m > n$, then there may be no solution, meaning no point at which all the constraints can be simultaneously satisfied. We can illustrate these points, as well as introduce an interesting model, in the “points rationing” example below.

Example S13.4 **Points Rationing**

Solve the constrained maximization problem

$$\max\ y = x_1^{0.5} x_2^{0.2} x_3^{0.2} \quad \text{s.t.} \quad \begin{cases} 100 - 2x_1 - 3x_2 = 0 \\ 20 - x_2 - 4x_3 = 0 \end{cases}$$

Solution

The Lagrange function is

$$\mathcal{L} = x_1^{0.5} x_2^{0.2} x_3^{0.2} + \lambda_1[100 - 2x_1 - 3x_2] + \lambda_2[20 - x_2 - 4x_3]$$

and the five first-order conditions are

$$\frac{\partial \mathcal{L}}{\partial x_1} = 0.5x_1^{-0.5} x_2^{0.2} x_3^{0.2} - 2\lambda_1 = 0$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = 0.2x_1^{0.5} x_2^{-0.8} x_3^{0.2} - 3\lambda_1 - \lambda_2 = 0$$

$$\frac{\partial \mathcal{L}}{\partial x_3} = 0.2x_1^{0.5} x_2^{0.2} x_3^{-0.8} - 4\lambda_2 = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = 100 - 2x_1 - 3x_2 = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = 20 - x_2 - 4x_3 = 0$$

which solve to give two potential sets of values for (x_1^*, x_2^*, x_3^*): (40.82, 6.12, 3.47) and (13.64, 24.24, −1.06). We investigate how to identify the true maximum in this type of problem in the section 13.2. For now, we can simply insert these alternative values into the objective function. Since the first set of solutions gives a positive value for the objective function while the second set gives a negative value, the solution we are looking for therefore appears to be (40.82, 6.12, 3.47). ■

S13.6 Constraints in Points Rationing

We take the case of a consumer with the same type of Cobb-Douglas utility function as before. We assume that the price of good 1 is $1, and that of good 2 is $2, and that the consumer has an income of $100. Now, however, we also assume that *points rationing* is in force. That is, the government specifies that when the consumer buys one unit of each good, she must hand over a specified number of ration coupons as well as the money price, and she is given an *initial endowment* of ration coupons. Assume that one unit of good 1 requires two coupons, and a unit of good 2 requires one coupon, and that she has in total 100 coupons. In effect the consumer faces *two* budget constraints: the *money-budget constraint*

$$1x_1 + 2x_2 = 100$$

and the *coupon-budget constraint*

$$2x_1 + 1x_2 = 100$$

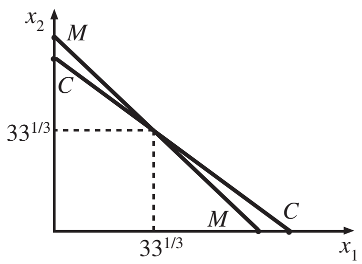

Figure S13.10 Constraints in the points-rationing problem

These are graphed in figure S13.10. The lines intersect at $x_1 = x_2 = 33.3$. The consumer's utility-maximization problem is now

$$\max u = x_1^\alpha x_2^\beta \quad \text{s.t.} \quad 1x_1 + 2x_2 = 100 \quad \text{and} \quad 2x_1 + 1x_2 = 100$$

We form the Lagrange function

$$\mathcal{L} = x_1^\alpha x_2^\beta + \lambda_1(100 - x_1 - 2x_2) + \lambda_2(100 - 2x_1 - x_2)$$

and maximizing this with respect to x_1, x_2, λ_1, and λ_2 gives

$$\frac{\partial \mathcal{L}}{\partial x_1} = \alpha x_1^{\alpha-1} x_2^\beta - \lambda_1 - 2\lambda_2 = 0$$
$$\frac{\partial \mathcal{L}}{\partial x_2} = \beta x_1^\alpha x_2^{\beta-1} - 2\lambda_1 - \lambda_2 = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = 100 - x_1 - 2x_2 = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = 100 - 2x_1 - x_2 = 0$$

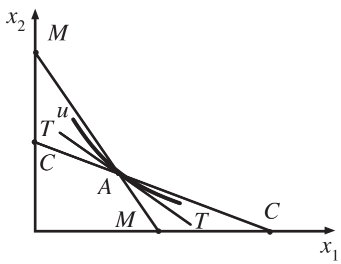

Figure S13.11 The marginal rate of substitution lies between the slopes of the constraints

However, before we start to solve these equations, we can save ourselves a little work by noting that the optimal values of x_1 and x_2 are determined by the last two conditions, the constraints of the problem, as we saw in figure S13.10. Since we have two constraints and two variables ($n = m$), this is a case in which the consumer has no choice—there is only one feasible solution, at $x_1 = x_2 = 33.3$. The first two conditions tell us that in general we do not have a tangency solution between an indifference curve and either one of the budget lines. To see this, refer first to figure S13.11. The solution is at point A, and at this point, the slope of the indifference curve, shown by the tangent line T, lies between the slope of the money-budget constraint, MM, and the slope of the coupon-budget constraint, CC.

(The alert reader may wonder about the case where the indifference curve just happens to be tangent to *one* of the budget lines at the kink. In that case, one of the Lagrange multipliers is zero, but full consideration of that case must await the Kuhn-Tucker conditions in chapter 15.)

Example S13.5 **Second-Order Conditions for the Land-Allocation Problem**

Confirm that the solution for the optimal land-allocation problem in example S13.2 is a true maximum.

Solution

The Hessian in this problem takes the form

$$H = \begin{bmatrix} -2.4l_1^{-1.4} & 0 & -1 \\ 0 & -1.28l_2^{-1.2} & -1 \\ -1 & -1 & 0 \end{bmatrix}$$

At the optimal solution, $l_1^* = 26.6$ and $l_2^* = 973.4$, this becomes

$$H^* = \begin{bmatrix} -0.0243 & 0 & -1 \\ 0 & -0.0003 & -1 \\ -1 & -1 & 0 \end{bmatrix}$$

Expanding the determinant along the first column gives

$$|H^*| = -0.0243(-1) - 1(-0.0003) = 0.0246 > 0$$

and so we have a true maximum. ■

Example S13.6 **Second-Order Conditions for the Cost-Minimization Problem**

Confirm that the solution to the constrained minimization problem in example S13.4 is a true minimum.

Solution

The Hessian matrix is

$$H = \begin{bmatrix} 0.25\lambda x_1^{-1.5} & 0 & -0.5x_1^{-0.5} \\ 0 & 0 & -1 \\ -0.5x_1^{-1.5} & -1 & 0 \end{bmatrix}$$

and so at the optimum $(x_1^*, x_2^*, \lambda^*) = (0.25, 0.5, 1)$,

$$H^* = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 0 & -1 \\ -1 & -1 & 0 \end{bmatrix}$$

Consequently $|H^*| = -(-1)[2(-1) - 0(1)] = -2 < 0$, and so we have a true minimum. ■

Example S13.7 **Second-Order Conditions with a Cobb-Douglas Production Function**

We return now to the example in which a firm with a Cobb-Douglas production function is seeking to minimize cost at a given level of output; that is, we have the problem

$$\min C = rK + wL \quad \text{s.t.} \quad y - K^a L^b = 0 \qquad a, b > 0$$

The Lagrange function is

$$\mathcal{L} = rK + wL + \lambda(y - K^a L^b)$$

We have already examined the first-order conditions at some length and so we proceed directly to the second-order conditions. We have the second-order partial derivatives

$$\begin{aligned} \mathcal{L}_{KK} &= -\lambda a(a-1)K^{a-2}L^b \\ \mathcal{L}_{LL} &= -\lambda b(b-1)K^a L^{b-2} \\ \mathcal{L}_{KL} = \mathcal{L}_{LK} &= -\lambda ab K^{b-1} L^{a-1} \end{aligned}$$

while the first-order partials of the constraint function in this case are

$$g_1 = -aK^{a-1}L^b, \quad g_2 = -bK^a L^{b-1}$$

Applying theorem 13.3, the first-order conditions characterize a true minimum if the Hessian

$$|H^*| = \begin{vmatrix} \mathcal{L}_{KK} & \mathcal{L}_{KL} & -aK^{a-1}L^b \\ \mathcal{L}_{LK} & \mathcal{L}_{LL} & -bK^a L^{b-1} \\ -aK^{a-1}L^b & -bK^a L^{b-1} & 0 \end{vmatrix} < 0$$

Expanding this determinant, substituting for the second-order partials of the Lagrange function, and rearranging eventually gives us the inequality

$$-2\lambda(ab)^2 K^{a-2}L^{b-2} < -\lambda[a^2 b(b-1)K^{a-2}L^{b-2} + b^2 a(a-1)K^{a-2}L^{b-2}]$$

So dividing through by $-\lambda(ab)^2 K^{a-2}L^{b-2}$ gives

$$0 > -(a+b)$$

which must be satisfied in this case because $a, b > 0$. Thus in the Cobb-Douglas case the first-order conditions give a true minimum.

An important point to note here is that the second-order condition was satisfied independently of the value of $a + b$. When $a + b < 1$, the Cobb-Douglas function is strictly concave, but when $a + b > 1$, it is *neither* concave *nor* convex, though it is still strictly quasiconcave. This example suggests therefore something we confirm in section 13.3: as a second-order condition, concavity, for a maximum, or convexity, for a minimum, are stronger than necessary; what really matters is the shape of the level curves of the function, namely whether the function is *quasiconcave* or *quasiconvex.* ■

Chapter S14 Comparative Statics

Contents

S14.1 Comparative Statics Examples

Example S14.1 **Government in the Simple Keynesian Model**

Suppose that there is a government in the simple Keynesian model. The government buys goods and services in the amount G and raises revenue through a tax on income. Let the rate of income tax be t, so that disposable income is $(1-t)Y$. If $t = 0.2$, $c = 0.8$, and $I = 500$, find the value of the government expenditure multiplier.

Solution

Aggregate demand is now $C + I + G$ and so in equilibrium

$$\begin{aligned} Y &= C + I + G \\ &= c(1-t)Y + I + G \\ &= 0.8(0.8)Y + 500 + G \end{aligned}$$

so that

$$Y^* = 1{,}400 + 2.8G$$

The government expenditure multiplier is

$$\frac{dY^*}{dG} = 2.8$$

■

Example S14.2 **Effects of a Change in Income on Price and Quantity**

Suppose that the demand and supply functions for a good are

$$D = 30 - 2p - y, \quad S = p$$

Find the effect of a change in income on equilibrium price and quantity.

Solution

In equilibrium, $D = S$, and so equilibrium price and quantity are

$$p^* = 10 - \frac{y}{3}, \quad q^* = 10 - \frac{y}{3}$$

and so the comparative statics we seek are

$$\frac{dp^*}{dy} = -\frac{1}{3}, \quad \frac{dq^*}{dy} = -\frac{1}{3}$$

The demand function indicates that this is an inferior good, with the consequence that equilibrium quantity and price both fall when income increases. ■

Example S14.3 **Effect of Tax on Monopoly Output**

Suppose that the monopolist has a demand curve $p = 180 - 2q$ and a total-cost function $C = q^2$. Find the effect of an increase in the output tax on equilibrium output.

Solution

Profit is $\pi = (180 - t)q - 3q^2$, and optimal output is

$$q^* = 30 - \frac{t}{6}$$

and the comparative-static effect is

$$\frac{dq^*}{dt} = -\frac{1}{6}$$

■

Example S14.4 Effect of a Change in the Discount Factor on Investment

In the planner's problem of choosing optimal investment, suppose that $Y_1^0 = 1{,}000$, and $\alpha = 0.5$. Solve for optimal investment as a function of the discount factor β and find the comparative-static effect of a change in β on I^*.

Solution

Making the appropriate substitutions into equation (14.7), we have

$$I^* = \frac{500\beta}{1 + 0.5\beta}$$

Using the quotient rule of differentiation (rule 8 in section 5.4), we have

$$\frac{dI^*}{d\beta} = \frac{500}{[1 + 0.5\beta]^2} > 0$$

So an increase in the discount factor, which places additional weight on future consumption in the utility function, increases investment. ■

Note from this example that comparative-statics exercises can be performed just as well to study the effect of a change in a model parameter on the endogenous variable, with the exogenous variable held constant.

Example S14.5 A Linear IS-LM Model

The expenditure function is $E = \bar{E} + 0.7Y - 100R$. The demand for money is $L = 30 + 0.2Y - 10.5R$. The money supply is $\bar{M}$. Solve for equilibrium Y and R as functions of autonomous investment $\bar{E}$ and the money supply $\bar{M}$. Show the effects on the equilibrium values of changes in $\bar{E}$ and $\bar{M}$.

Solution

The equilibrium conditions are

$$Y - (0.7Y - 100R) - \bar{E} = 0$$
$$30 + 0.2Y - 10.5R - \bar{M} = 0$$

Gathering terms, we can write these as the linear system

$$\begin{bmatrix} 0.3 & 100 \\ 0.2 & -10.5 \end{bmatrix} \begin{bmatrix} Y \\ R \end{bmatrix} = \begin{bmatrix} \bar{E} \\ \bar{M} - 30 \end{bmatrix}$$

The determinant of the square matrix is $-10.5(0.3) - 100(0.2) = -23.15$. Using Cramer's rule, we have as solutions

$$Y^* = \frac{\begin{vmatrix} \bar{E} & 100 \\ \bar{M} - 30 & -10.5 \end{vmatrix}}{-23.15} = \frac{-(10.5\bar{E} + 100\bar{M} - 3{,}000)}{-23.15}$$
$$= 0.45\bar{E} + 4.32M^0 - 129.59$$

$$R^* = \frac{\begin{vmatrix} 0.3 & \bar{E} \\ 0.2 & \bar{M} - 30 \end{vmatrix}}{-23.15} = \frac{0.3\bar{M} - 0.2\bar{E} - 9}{-23.15}$$
$$= 0.39 - 0.013\bar{M} + 0.009\bar{E}$$

It follows that

$$\frac{\partial Y^*}{\partial \bar{E}} = 0.45 \quad \frac{\partial Y^*}{\partial \bar{M}} = 4.32 \quad \frac{\partial R^*}{\partial \bar{E}} = 0.009 \quad \frac{\partial R^*}{\partial \bar{M}} = -0.013$$

■

Example S14.6 **Effects of a Wage Change in a Cobb-Douglas Model**

Use the Cobb-Douglas production function $y = pL^{\alpha}K^{\beta}$ to find the comparative-static effects $\partial L^*/\partial w$ and $\partial K^*/\partial w$.

Solution

The first-order and second-order conditions for this were derived in the example in section 12.2, and the Hessian is

$$|D| = p^2\alpha\beta L^{2\alpha-2}K^{2\beta-2}(1 - \alpha - \beta) > 0$$

for $1 > \alpha + \beta$. Since $f_{KK} = \beta(\beta - 1)L^{\alpha}K^{\beta-2} < 0$, and $f_{KL} = \alpha\beta L^{\alpha-1}K^{\beta-1} > 0$, we have

$$\frac{\partial L^*}{\partial w} = p\frac{\beta(\beta - 1)L^\alpha K^{\beta-2}}{|D|} < 0$$

$$\frac{\partial K^*}{\partial w} = -p\frac{\alpha\beta L^{\alpha-1} K^{\beta-1}}{|D|} < 0$$

Clearly, in the case of a Cobb-Douglas production function, an increase in the wage unambiguously reduces the demand for capital. ■

Example S14.7 **The Slutsky Equation for Cobb-Douglas Preferences**

Find the Slutsky equation for x_1^* when the consumer has Cobb-Douglas preferences $u(x_1, x_2) = x_1^{0.5}x_2^{0.5}$.

Solution

We have

$$u_{11} = -0.25x_1^{-1.5}x_2^{0.5} < 0, \quad u_{22} = -0.25x_1^{0.5}x_2^{-1.5} < 0$$
$$u_{12} = u_{21} = 0.25x_1^{-0.5}x_2^{-0.5} > 0$$

and so we have, using equation (14.21) and the fact that $\lambda^* = u_1/p_1$,

$$\frac{\partial x_1^*}{\partial p_1} = -\frac{0.5x_1^{-0.5}x_2^{0.5}p_2^2}{p_1|D|} + x_1^*\frac{-p_1 0.25x_1^{0.5}x_2^{-1.5} - p_2 0.25x_1^{-0.5}x_2^{-0.5}}{|D|} < 0$$

So, in the case of Cobb-Douglas preferences, since x_1^* is a normal good, an increase in p_1 surely reduces the demand x_1^*. ■

S14.2 The Profit Function

A competitive firm wishes to maximize its profit

$$\pi = py - wL - rK$$

where y is output, L is labor, and K is capital, with p, w, and r, their respective prices. It has the production function $y = f(L, K)$ with the usual properties and, in particular, decreasing returns to scale. The Lagrange function is

$$\mathcal{L} = py - wL - rK + \lambda[f(L, K) - y]$$

The first-order conditions are

$$\begin{aligned} p - \lambda &= 0 \\ \lambda f_L(L, K) - w &= 0 \\ \lambda f_K(L, K) - r &= 0 \\ -y + f(L, K) &= 0 \end{aligned}$$

These give solutions for the input-demand functions $L(p, w, r)$ and $K(p, w, r)$, and the *output-supply function* $y(p, w, r)$. Inserting these functions into the profit equation gives the *profit function*

$$\pi = py(p, w, r) - wL(p, w, r) - rK(p, w, r) = V(p, w, r)$$

which is the value function in this problem. Now notice that the exogenous variables p, w, and r appear only in the objective function and so applying the envelope theorem gives

$$\begin{aligned} \frac{\partial V}{\partial p} &= \frac{\partial \mathcal{L}}{\partial p} = y(p, w, r) \\ \frac{\partial V}{\partial w} &= \frac{\partial \mathcal{L}}{\partial w} = -L(p, w, r) \\ \frac{\partial V}{\partial r} &= \frac{\partial \mathcal{L}}{\partial r} = -K(p, w, r) \end{aligned}$$

That is, differentiating the profit function with respect to the prices gives the output-supply and input-demand functions for the firm. This is often referred to as **Hotelling's lemma.**

Example S14.8 **Profit Function for a Firm**

Find the profit function for a competitive firm with the production function

$$y = L^{0.4} K^{0.4}$$

Solution

The first-order conditions for the profit-maximization problem are

$$\begin{aligned} p - \lambda &= 0 \\ \lambda 0.4 L^{-0.6} K^{0.4} - w &= 0 \end{aligned}$$

$$\lambda 0.4 L^{0.4} K^{-0.6} - r = 0$$
$$y = L^{0.4} K^{0.4}$$

Substitute p for λ in the second and third conditions, and then note that we have a linear system in logarithms

$$\begin{bmatrix} -0.6 & 0.4 & 0 \\ 0.4 & -0.6 & 0 \\ 0.4 & 0.4 & -1 \end{bmatrix} \begin{bmatrix} \hat{L} \\ \hat{K} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} \hat{w} - \hat{p} - \log 0.4 \\ \hat{r} - \hat{p} - \log 0.4 \\ 0 \end{bmatrix}$$

where a "hat" over a variable denotes its log. Solving this system and taking antilogs gives the input-demand and output-supply functions

$$L = 0.01 w^{-3} r^{-2} p^5$$
$$K = 0.01 w^{-2} r^{-3} p^5$$
$$y = 0.03 w^{-2} r^{-2} p^4$$

To obtain the profit function, we substitute into the expression

$$\pi = px - wl - rk$$

to get

$$\pi = 0.01 w^{-2} r^{-2} p^5$$

which is the value function—the profit function—in this example. Figure S14.1 illustrates the shape of this function. Notice that it is convex in w, r, and p. This can be shown to be a general property of the profit function. ■

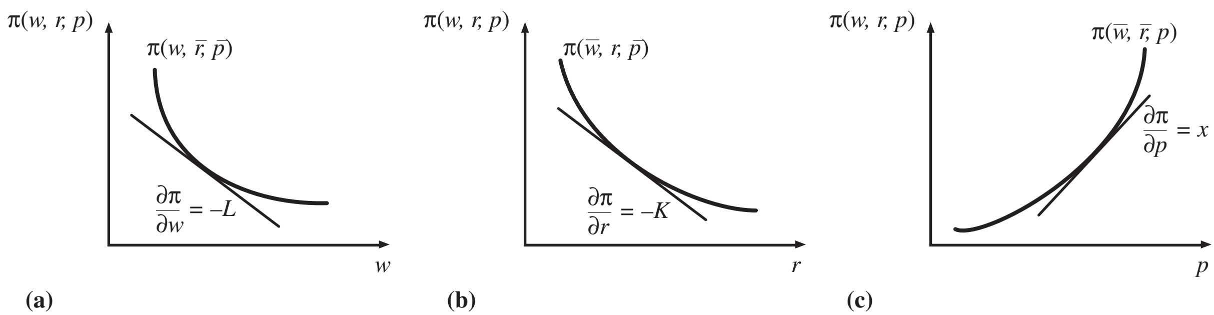

Figure S14.1 Cross sections through the profit function in the Cobb-Douglas case

S14.3 The Indirect Utility Function

Consider the standard consumer problem

$$\max u(x_1, x_2) \quad \text{s.t.} \quad p_1x_1 + p_2x_2 = m$$

From the first-order conditions we obtain the demand functions $x_1(p_1, p_2, m)$, and $x_2(p_1, p_2, m)$, and substituting these into the utility function gives

$$u = u(x_1(p_1, p_2, m), x_2(p_1, p_2, m)) = V(p_1, p_2, m)$$

which is the value function in this problem—the *indirect utility function*. Applying the envelope theorem gives

$$\frac{\partial V}{\partial m} = \lambda^*$$

It then seems reasonable to call λ the *marginal utility of income* and of course since it is an endogenous variable, it is also a function of prices and income in general. Next, notice that prices appear only in the constraint of the problem, and applying the envelope theorem again gives

$$\frac{\partial V}{\partial p_i} = \frac{\partial \mathcal{L}}{\partial p_i} = -\lambda^* x_i(p_1, p_2, m), \qquad i = 1, 2$$

This result, known as **Roy's identity,** tells us that, given that income has positive marginal utility, a change in the price of a good has a negative effect on utility which is proportional to the amount of the good consumed.

S14.4 The Expenditure Function

Now consider the *dual problem* to the indirect utility function. Instead of maximizing utility subject to a budget constraint, we *minimize expenditure* subject to a *utility constraint*. The idea here is that the utility constraint defines a particular standard of living and so we are trying to find the cheapest way to achieve a particular standard of living. The problem is

$$\min_{x_1, x_2} E = p_1x_1 + p_2x_2 \quad \text{s.t.} \quad u(x_1, x_2) = \bar{u}$$

where E denotes expenditure and $\bar{u}$ is the given utility level. The Lagrange function is

$$\mathcal{L} = p_1x_1 + p_2x_2 + \mu[\bar{u} - u(x_1, x_2)]$$

where μ is the Lagrange multiplier. The solution to this includes two demand functions, $x_1(p_1, p_2, u)$ and $x_2(p_1, p_2, u)$, which must differ from those derived in the previous example since they have u as an argument, not m. Substituting into the objective function gives

$$E = p_1 x_1(p_1, p_2, u) + p_2 x_2(p_1, p_2, u) = e(p_1, p_2, u)$$

where e is called the **expenditure function** and is the value function in this problem. Since u is in the constraint, we see that

$$\frac{\partial E}{\partial u} = \mu^*$$

and we can think of μ as the *marginal cost of utility*. Next note that prices enter only into the objective function and so the envelope theorem gives

$$\frac{\partial E}{\partial p_i} = \frac{\partial \mathcal{L}}{\partial p_i} = x_i(p_1, p_2, u), \qquad i = 1, 2$$

This is known as **Shephard's lemma.** If the consumer is consuming x_i units of good i and its price increases by, say, one cent, then to maintain the same standard of living, defined as a particular utility level, the consumer requires approximately x_i cents more income.

Example S14.9 Expenditure Function for a Consumer

A consumer has the utility function

$$u = x_1^a x_2^b, \qquad a, b > 0;\ a + b = 1$$

We solved this problem in chapter 12, and know the demand functions to be

$$x_1 = \frac{am}{p_1}, \quad x_2 = \frac{bm}{p_2}$$

Substituting these values into the utility function gives the indirect utility function

$$u = \left(\frac{am}{p_1}\right)^a \left(\frac{bm}{p_2}\right)^b = a^a b^b p_1^{-a} p_2^{-b} m$$

As we expected, it is decreasing in prices and increasing in income. It gives the maximum utility level achievable at prices p_1, p_2, and income m. Inverting the function gives

$$m = a^{-a} b^{-b} p_1^a p_2^b u$$

which is increasing in prices and utility. We can interpret this function as giving the income required to achieve utility u at prices p_1 and p_2 when the consumer chooses optimally. But, this is simply the definition of the expenditure function, so is the expenditure function in this case. To confirm this, note that the solution to the problem

$$\min\ p_1 x_1 + p_2 x_2 \quad \text{s.t.} \quad u = x_1^a x_2^b$$

gives the demand functions

$$x_1 = a^b b^{-b} p_1^{-b} p_2^b u, \quad x_2 = a^{-a} b^a p_1^a p_2^{-a} u$$

Substituting these into $m = p_1 x_1 + p_2 x_2$ gives

$$m = a^b b^{-b} p_1^{1-b} p_2^b u + a^{-a} b^a p_1^a p_2^{1-a} u$$

and then using $a + b = 1$, we find

$$m = a^{-a} b^{-b} p_1^a p_2^b u$$

as before. Note finally that the expenditure function in this case is strictly concave in prices, since

$$\frac{\partial^2 m}{\partial p_1^2} = -(1-b) a^b b^{1-b} p_1^{-(1+b)} p_2^b u < 0$$

and similarly for p_2. Concavity (though not necessarily strict concavity) is a feature of expenditure functions in general. This is illustrated in figure S14.2. ■

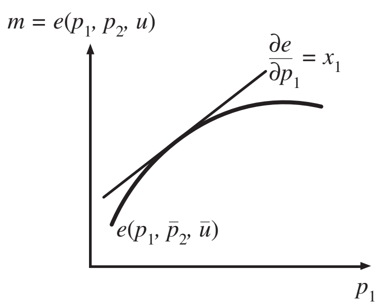

Figure S14.2 The expenditure function is concave in prices

Chapter S15 Concave Programming and the Kuhn-Tucker Conditions

Contents

S15.1 Cost-Minimization

The firm's cost-minimization problem is

$$\min C = wL + rK \quad \text{s.t.} \quad f(L, K) - y \geq 0 \quad \text{and} \quad L \geq 0, K \geq 0$$

Since minimizing a function is equivalent to maximizing the negative of that function, we can put the problem in the standard form by taking $-(wL + rK)$ as our maximand. Then the Lagrange function is

$$\mathcal{L} = -(wL + rK) + \lambda[f(L, K) - y]$$

and the K-T conditions are

$$\frac{\partial \mathcal{L}}{\partial L} = -w + \lambda^* f_L(L^*, K^*) \leq 0, \qquad L^* \geq 0;\, L^*(\lambda^* f_L - w) = 0$$
$$\frac{\partial \mathcal{L}}{\partial K} = -r + \lambda^* f_K(L^*, K^*) \leq 0, \qquad K^* \geq 0;\, K^*(\lambda^* f_K - r) = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = f(L^*, K^*) - y \geq 0, \qquad \lambda^* \geq 0;\, \lambda^*(f(L^*, K^*) - y) = 0$$

We begin by making the assumption that both inputs are essential if any output is to be produced, so that we are only interested in interior solutions with both inputs strictly positive. It follows that the first two conditions have to be equalities. Since the input prices w and r are both positive, so must be λ^*, and so from the third condition we have that the constraint must be binding. This tells us that the firm will not produce inefficiently: if $y < f(L^*, K^*)$, then more output could be produced with the same input quantities and this would be inefficient. If the constraint is binding, no more output can be produced with the same inputs.

In this case, then, the optimum is exactly the one we found earlier by using the standard Lagrangean approach. The advantage of using the K-T approach here is that it clarifies the conditions under which that solution is obtained and also facilitates analysis of more general cases, for example, if output can be produced with only one input or if one input price were zero (see question 2 of the exercises).

S15.2 The Linear-Programming Problem

The earliest form of concave-programming problem studied by mathematicians and economists was the linear-programming problem, in which the objective and constraint functions are all linear. The significance of this was that relatively straightforward methods could be used to solve numerical problems of this type, something rather harder to do in the case of nonlinear concave-programming problems. But in addition certain interesting theoretical ideas emerged that attracted a lot of attention in economics. In this example, we will develop some of these ideas, simply by treating a linear-programming problem as a concave programming problem and applying the Kuhn-Tucker theorem to it.

To get started, we take a specific problem. A firm uses three inputs—labor L, machine time M, and raw materials R—to produce two outputs, x_1 and x_2. It has fixed amounts of these inputs available, L^0, M^0, and R^0. The selling price of output $i = 1, 2$ is $p_i > 0$, and one unit of output i requires l_i units of labor, m_i units of machine time, and r_i units of raw material to produce, where all these parameters are independent of the scale of output. The profit-maximizing problem of the firm is then the special form of concave-programming problem:

$$\max\ p_1x_1 + p_2x_2 \quad \text{s.t.} \quad \begin{cases} l_1x_1 + l_2x_2 \le L^0 \\ m_1x_1 + m_2x_2 \le M^0 \\ r_1x_1 + r_2x_2 \le R^0 \\ x_1, x_2 \ge 0 \end{cases}$$

Maximization of revenue is equivalent to maximizing profit because we assume that input costs are all fixed. The constraints simply say that a particular output pair (x_1, x_2) is feasible if and only if it is nonnegative and does not require more than the available amounts of inputs to be produced. We assume the feasible set defined by these constraints is nonempty.

The Lagrange function for the problem is

$$\begin{aligned} \mathcal{L} = p_1x_1 + p_2x_2 &+ \lambda(L^0 - l_1x_1 - l_2x_2) \\ &+ \mu(M^0 - m_1x_1 - m_2x_2) \\ &+ \rho(R^0 - r_1x_1 - r_2x_2) \end{aligned}$$

The K-T conditions are

$$
\begin{aligned}
p_i - \lambda^* l_i - \mu^* m_i - \rho^* r_i \leq 0, &\quad x_i^* \geq 0, \quad x_i^*(p_i - \lambda^* l_i - \mu^* m_i - \rho^* r_i) = 0 \\
L^0 - l_1 x_1^* - l_2 x_2^* \geq 0, &\quad \lambda^* \geq 0, \quad \lambda^*(L^0 - l_1 x_1^* - l_2 x_2^*) = 0 \\
M^0 - m_1 x_1^* - m_2 x_2^* \geq 0, &\quad \mu^* \geq 0, \quad \mu^*(M^0 - m_1 x_1^* - m_2 x_2^*) = 0 \\
R^0 - r_1 x_1^* - r_2 x_2^* \geq 0, &\quad \rho^* \geq 0, \quad \rho^*(R^0 - r_1 x_1^* - r_2^*) = 0
\end{aligned}
$$

The Lagrange multipliers are ususally referred to as *dual variables* in linear programming. The key point is their interpretation as the **shadow prices** of the input constraints. At the optimal solution, the value of λ^*, μ^*, or ρ^* gives the increase in revenue the firm would earn if it acquired a little bit more of the respective inputs and allocated that optimally between the outputs. The last three conditions also tell us that if a shadow price is positive, all of the corresponding input is used up, while if at the optimum, there is some amount of an input unused; its shadow price is zero.

It is always useful to use a diagram to obtain a sense of the solution possibilities. In figure S15.1 we see that there may be as many as seven. The lines denoted R are iso-revenue lines; this means that they show (x_1, x_2)-pairs that generate the same revenue, and so they satisfy

$$
R = p_1 x_1 + p_2 x_2
$$

for given R. They therefore have slope $-p_1/p_2$, and at given prices, the higher the line, the greater the revenue. The shaded area is the feasible set in each case, and we have assumed that the exact form of the constraints is such that each *could* be binding. If one constraint coincided with another or lay entirely outside another, then it could be dropped and the number of solution possibilities would consequently fall. The constraint L^0 has slope $-l_1/l_2$, M^0 has slope $-m_1/m_2$, and R^0 has slope $-r_1/r_2$. No point above any constraint line is feasible. The feasible set is the intersection of the sets of points lying on or below a constraint line. It is assumed in the figure that

$$
\frac{l_1}{l_2} < \frac{m_1}{m_2} < \frac{r_1}{r_2}
$$

Then, which of the solution possibilities results depends on the value of p_1/p_2 relative to the slopes of the constraint lines.

In figure S15.1 (a), we have $p_1/p_2 < l_1/l_2$, and so the solution is at point a, with $x_2^* > 0$ and $x_1^* = 0$. Note also that only the labor constraint is binding, so $\lambda^* > 0$ while $\mu^* = \rho^* = 0$.

In figure S15.1 (b), we have $p_1/p_2 = l_1/l_2$. The highest possible revenue line coincides with the labor constraint. In that case any point on the segment $[a, b]$ is optimal. The labor constraint is binding, so $\lambda^* > 0$. However, note that *small*

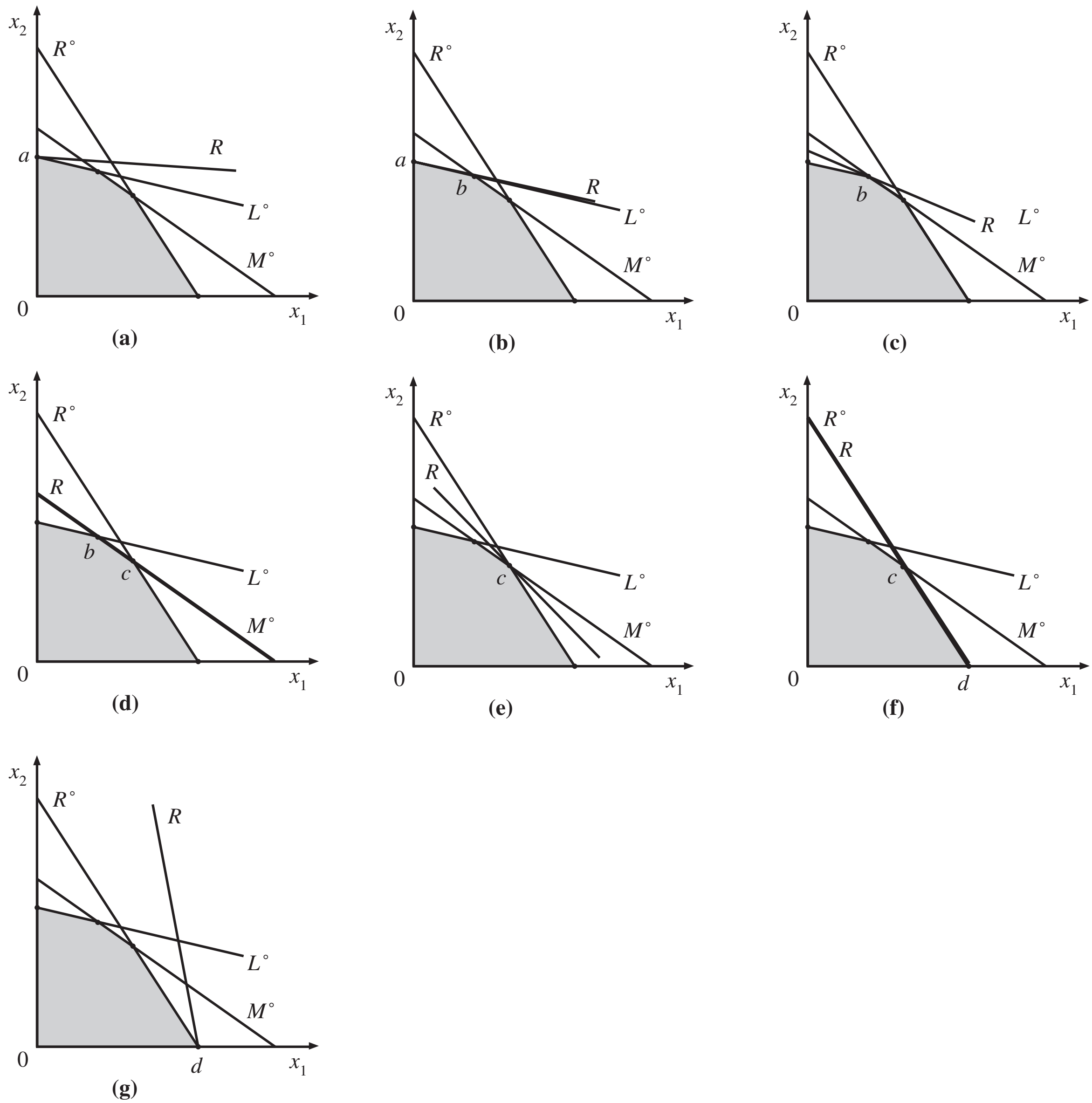

Figure S15.1 Solution possibilities in the linear-programming problem variations in either of the two other constraints cannot change the maximized value of revenue, and so again $\mu^* = \rho^* = 0$.

In figure S15.1 (c), we have $l_1/l_2 < p_1/p_2 < m_1/m_2$. There is a unique optimum at b, with $x_1^* > 0$, $x_2^* > 0$. In addition both labor and machine-time constraints are binding, and so $\lambda^* > 0$, $\mu^* > 0$ while $\rho^* = 0$.

In figure S15.1 (d), we have $p_1/p_2 = m_1/m_2$. Any point on the segment $[b, c]$ is optimal, and so $x_1^* > 0$, $x_2^* > 0$. The machine-time constraint is certainly binding, and so $\mu^* > 0$. However, small variations in the other two constraints leave maximum revenue unchanged, and so $\lambda^* = \rho^* = 0$.

In figure S15.1 (e), we have $m_1/m_2 < p_1/p_2 < r_1/r_2$, and there is a unique solution at c with $x_1^* > 0$, $x_2^* > 0$. The machine-time and raw-material constraints are both binding, and so $\mu^* > 0$, $\rho^* > 0$ but $\lambda^* = 0$.

In figure S15.1 (f), we have $p_1/p_2 = r_1/r_2$, and so any point on the segment $[c, d]$ is optimal. The raw-material constraint is certainly binding, and so $\rho^* > 0$, but small variations in the other two constraints leave maximized revenue unchanged, and so $\lambda^* = \mu^* = 0$.

In figure S15.1 (g), we have $p_1/p_2 > r_1/r_2$, and so there is a unique solution with $x_1^* > 0$, $x_2^* = 0$. Only the raw-material constraint is binding, and so $\rho^* > 0$ and $\lambda^* = \mu^* = 0$.

One notable feature of the solutions is that the optimum could always be taken to be at a *corner point* of the upper boundary of the feasible set, a point such as a, b, c, or d. This is what greatly facilitates numerical solution of linear problems: it is necessary simply to evaluate the objective function at corner points, rather than over the entire feasible set. However, here we are interested in the economic, rather than the computational, aspects of the solution.

A second notable feature is that there were never more than two binding constraints at any solution: at least one constraint was always nonbinding. This is a consequence of the fact that there are two variables in the problem. Thus consider the last three K-T conditions above and note that if at most two variables can be strictly positive, then three constraints cannot be nontrivially binding, because that would give three equations in only two unknowns.

Consider now the first of the K-T conditions, which relates to the optimal outputs. Suppose that both $x_1^* > 0$ and $x_2^* > 0$, and that only the raw-material constraint is nonbinding so that we have $\rho^* = 0$. Then these conditions take the form

$$p_1 - \lambda^* l_1 - \mu^* m_1 = 0$$
$$p_2 - \lambda^* l_2 - \mu^* m_2 = 0$$

Recall that λ^* and μ^* are interpreted as the shadow prices of the inputs. Then these conditions have the interpretation of "marginal revenue = marginal cost" conditions. To see this, note that l_i is the amount of labor used in producing one unit of x_i so that $\lambda^* l_i$ is the *imputed cost,* valued at the shadow price of labor, of the amount of labor used to produce one unit of x_i. Likewise m_i is the amount of machine time used per unit of x_i, and so $\mu^* m_i$ is the *imputed cost* of the machine time used per unit of x_i. Therefore $\lambda^* l_i + \mu^* m_i$ is the imputed cost of a unit of x_i, while p_i is the marginal revenue of x_i.

Notice that whatever may be the actual prices the firm pays for the inputs, these unit costs are evaluated at the shadow prices because these are the appropriate measures of the *marginal opportunity costs* of the inputs to the firm. In particular, note that the cost of raw material plays no part in the unit cost calculation. Because $\rho^* = 0$, this input is not *relatively scarce* to the firm.

Returning to the K-T conditions, we see that

$$p_i - \lambda^* l_i - \mu^* m_i < 0$$

implies that $x_i^* = 0$. This says that if the marginal revenue falls short of its unit cost *evaluated at the appropriate shadow prices,* then the good should not be produced. All the resources should be allocated to the other good for which the equality in this condition will hold.

Finally, note that multiplying through the first condition by x_1^*, and the second by x_2^*, and adding gives

$$\begin{aligned} p_1 x_1^* + p_2 x_2^* &= \lambda^*(l_1 x_1^* + l_2 x_2^*) + \mu^*(m_1 x_1^* + m_2 x_2^*) \\ &= \lambda^* L^0 + \mu^* M^0 \end{aligned}$$

Therefore, not only do the shadow prices give the marginal value or marginal-opportunity cost of each input, but they do so in such a way that the entire revenue of the firm is imputed to the inputs: we could regard $\lambda^* L^0$ as the share of revenue that can be imputed to labor, and $\mu^* M^0$ the share imputed to machine time, and these shares exactly exhaust available revenue.

A slight paradox here is that this makes it seem as though raw materials are valueless to the firm, which is clearly not true in an absolute sense, because raw material is used (in the amount $r_1 x_1^* + r_2^* x_2^* < R^0$) and output could not be produced without it. The point is that *at the margin* the stock of raw material is valueless, because the firm has more of it than is optimal to use. The result above then shows that the sum of the costs of the inputs evaluated at their shadow prices, just equals the total revenue. This is essentially a result of the constant-returns-to-scale assumption underlying the linear model.

Chapter S16 Integration

Contents

S16.1 More on Consumer Surplus Measurement

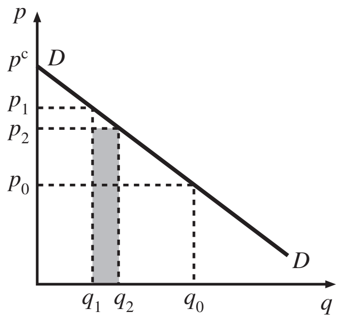

Figure S16.1 Approximation of a consumer's willingness-to-pay for a change in the price from p_1 to p_2

The term $\int_0^{q_0} D^{-1}(q)\,dq$ is sometimes referred to as the **gross surplus** of the consumer or an approximation of the willingness to pay for q_0 units of output. To see why we use this term, consider any price greater than p_0 but low enough that the consumer will purchase some positive amount q, where $0 < q < q_0$, say, $p = p_1$ and $q = q_1$ in figure S16.1.

If the price were to fall slightly to p_2 the consumer would be willing to buy $q_2 - q_1$ more units, or, in other words, the consumer would be willing to pay the amount shown in the shaded area to obtain the additional $q_2 - q_1$ units, having already been consuming q_1 units. The consumer would not have been willing to pay more than this since, if the price had fallen to a value greater than p_2, less would have been purchased. By considering a series of such price reductions, beginning at price $p = p^c$, where $q = 0$, to price $p = p_0$, and letting the size of each price reduction approach zero (in the limit), we see that we have simply generated the Riemann integral

$$\int_0^{q_0} D^{-1}(q)\,dq$$

and this represents the maximum willingness to pay for $q = q_0$ units of the good. Since $p_0 q_0$ is the amount actually paid, the difference, given in definition 16.5, is the *surplus* for the consumer.

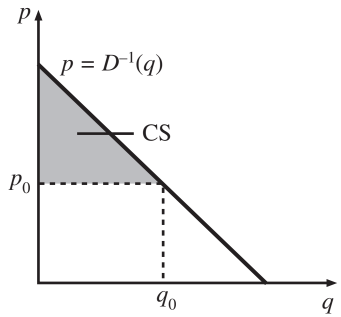

Figure S16.2 Consumer surplus (CS) for a consumer at price $p = p_0$

There is one flaw in this argument. The point (q_0, p_0) on the demand schedule reflects the optimal consumption decision of the consumer, who pays price p_0 for each marginal unit from $q = 0$ to $q = q_0$, not a series of prices all of which exceed p_0 except for the price of the final marginal unit. In the latter case the consumer's income net of expenditures on this good is less than in the case where the single price $p = p_0$ is charged for all units purchased, and so unless there is no income effect on demand for this good, there would be a different consumption decision taken in each of these cases. Therefore we should think of consumer surplus only as an approximation of the benefit of consuming a good at a certain price.

The diagrammatic interpretation of consumer surplus is shown in figure S16.2. It is the area below the demand curve and above price over the interval $q \in [0, q_0]$.

Example S16.1

For a consumer with demand function $q = 50 - 2p$, find

(i) CS at price $p_0 = 20$
(ii) CS at price $\hat{p} = 15$
(iii) ΔCS (change in consumer surplus) from price change $p_0 = 20$ to $\hat{p} = 15$

Illustrate each answer in a graph and show how to compute ΔCS using the demand function $q = D(p)$, rather than using the *inverse*-demand function.

Solution

(i) At $p_0 = 20$, $q_0 = 10$. The inverse-demand function is $p = 25 - q/2$ and so

$$\begin{aligned}
\mathrm{CS}(p_0 = 20) &= \int_0^{10} \left(\frac{25 - q}{2}\right) dq - 20(10) \\
&= \left[\frac{25q - q^2}{4}\right]_0^{10} - 200 \\
&= \left[\left(\frac{250 - 100}{4}\right) - (0 - 0)\right] - 200 = 25
\end{aligned}$$

This corresponds to the area indicated in figure S16.3 (a).

(ii) At $\hat{p} = 15$, $\hat{q} = 20$, and so

$$\begin{aligned}
\mathrm{CS}(\hat{p} = 15) &= \int_0^{20} \left(\frac{25 - q}{2}\right) dq - 15(20) \\
&= \left[\frac{25q - q^2}{4}\right]_0^{20} - 300 \\
&= \left[\left(\frac{500 - 400}{4}\right) - (0 - 0)\right] - 300 = 100
\end{aligned}$$

This corresponds to the area indicated in figure S16.3 (b).

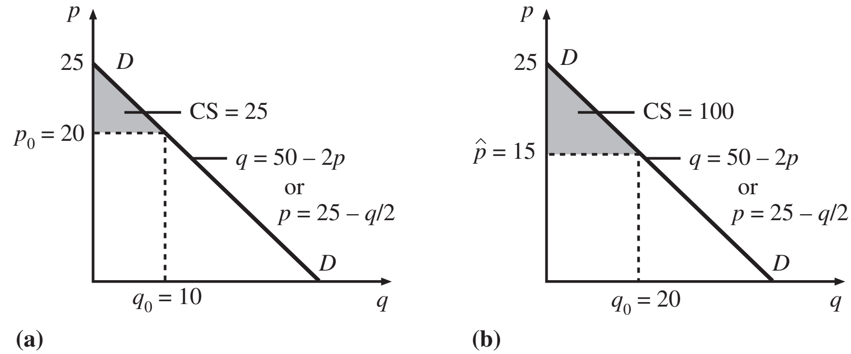

Figure S16.3 Computation of consumer surplus at prices $p = 20$ and $p = 15$ (example S16.1)

(iii) From (i) and (ii) we can see that changing price from $p_0 = 20$ to $\hat{p} = 15$ leads to an increase in consumer surplus of amount $100 - 25 = 75$. Noting that the inverse of the inverse-demand function is simply the demand function

$$q = D(p) = 50 - 2p$$

we can compute the change in consumer surplus according to the formula

$$\begin{aligned}\Delta\text{CS} &= \int_{\hat{p}}^{p_0} D(p)\,dp \\ &= \int_{15}^{20} (50 - 2p)\,dp \\ &= [50p - p^2]_{15}^{20} = 75\end{aligned}$$

The same rationale given in example 16.4 for using the inverse function to find the change in producer surplus applies here, as is illustrated by figure S16.4 where we see that area $ABCD$ in figure S16.4 (a) corresponds to area $abcd$ in figure S16.4 (b).

■

To indicate the effect of a price increase from p_0 to $\hat{p}$, where $\hat{p} > p_0$, we write

$$\Delta\text{CS} = \int_{\hat{p}}^{p_0} D(p)\,dp = -\int_{p_0}^{\hat{p}} D(p)\,dp$$

Notice that in evaluating ΔCS it is conventional to use the *original* price (p_0) as the upper limit for the integral and the *new* price ($\hat{p}$) as the lower limit, so that ΔCS is positive for a price fall and negative for a price rise.

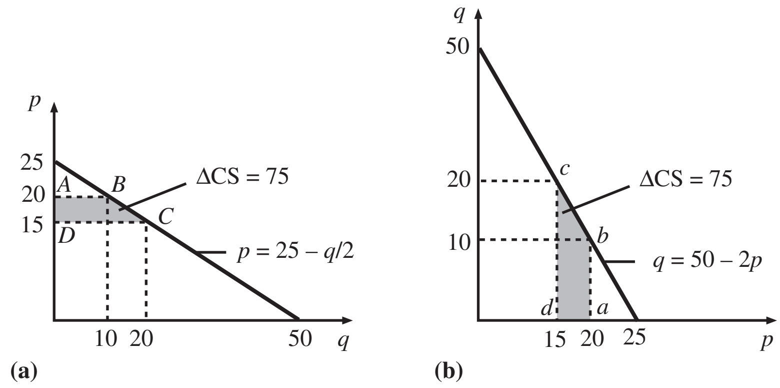

Figure S16.4 Two methods of computing the change in consumer surplus (example S16.1 (iii))

We can also use the consumer surplus concept to measure the impact of a product becoming unavailable. This could happen, for example, as a result of an import ban. To deal with this issue, we start by defining the *choke price* of a good, $p = p_c$, as the minimum price for which demand becomes zero. That is, demand is *choked off*. For example, the choke price for the demand function $q = 50 - 2p$ is $p_c = 25$ (see figure S16.3). The impact on consumers of a product becoming unavailable is equivalent to an increase in its price from its existing price, p_0, to the choke price.

Example S16.2 The demand function for a product is $q = 15 - 3p^{1/2}$. Find the loss of consumer surplus resulting from a ban on purchases of this product if its current price is $p = 9$.

Solution

The choke price solves for $q = 0$ and so $15 - 3p_c^{1/2} = 0$ or $p_c = 25$. Therefore the change in consumer surplus resulting from the ban is

$$\begin{aligned}
\Delta\text{CS} &= \int_{25}^{9} (15 - 3p^{1/2})\, dp \\
&= -\int_{9}^{25} (15 - 3p^{1/2})\, dp \\
&= -[15p - 2p^{3/2}]_9^{25} \\
&= -[375 - 2(125)] + [135 - 2(27)] = -44
\end{aligned}$$

The result is illustrated in figure S16.5. ■

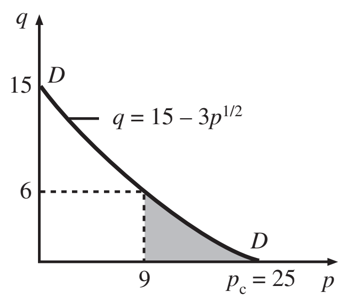

Figure S16.5 Loss of consumer surplus resulting from the withdrawal of a product from the market (example S16.2)

Chapter S18 Linear, First-Order Difference Equations

Contents

S18.1 A Modified Cobweb Model

In the cobweb model of price determination that was examined in section 18.1, modify the way in which price expectations are formed, as follows:

$$E_{t-1}(p_t) = p_{t-1} + \theta(\bar{p} - p_{t-1}), \qquad 0 \leq \theta \leq 1$$

We assume that suppliers have an accurate forecast of the steady-state equilibrium price, $\bar{p}$. As a result, in period $t-1$, the next period price is expected to equal current price (p_{t-1}), plus a fraction, θ, of the difference between the steady-state price and the current price. If the current price is below the steady-state price, price is expected to rise; if the current price is above the steady-state price, price is expected to fall. If $\theta = 0$, this model reduces to the basic cobweb model examined in section 18.1 of the textbook. If $\theta = 1$, suppliers expect price to adjust to the equilibrium price in one period.

Example S18.1

Solve the difference equation for price and analyze the convergence properties of the solution.

Solution

Substituting the modified price expectation equation into the supply function, and then setting quantity supplied equal to quantity demanded gives

$$A + Bp_t = F + G(1-\theta)P_{t-1} + \theta G\bar{p}$$

Solving for p_t gives

$$p_t = (1-\theta)\frac{G}{B}p_{t-1} + \frac{F - A + \theta G\bar{p}}{B}$$

To find the steady-state price, set $p_t = p_{t-1} = \bar{p}$. After some manipulation this gives

$$\bar{p} = \frac{A - F}{G - B}$$

which is the same steady-state equilibrium price as in the basic cobweb model. By theorem 18.1, the solution to the difference equation is

$$p_t = \left((1-\theta)\frac{G}{B}\right)^t p_0 + \frac{F - A + \theta G\bar{p}}{B}\left[\frac{1 - [(1-\theta)G/B]^t}{1 - (1-\theta)G/B}\right]$$

Simplifying, rearranging, and using the fact that $(F - A) = (B - G)\bar{p}$, gives

$$p_t = (p_0 - \bar{p})\left((1-\theta)\frac{G}{B}\right)^t + \bar{p} \tag{S18.1}$$

Like the basic cobweb model, price in this model oscillates whether it converges or not because $(1-\theta)G/B$ is negative for the usual case of negatively sloping demand and positively sloping supply. However, this modified model is more likely to satisfy the convergence condition than the basic cobweb model because the absolute value of $(1-\theta)G/B$ is more likely to be less than 1 than is $G/$B. For example, even if the absolute value of G is 5 times as large as B, price converges to its equilibrium value if θ is larger than 0.8. ■

S18.2 A Partial Adjustment Model of Energy Demand

Suppose that the desired long-run industrial demand for energy is a function of the price of energy, the prices of other factors of production, and the price of industrial output. We express this as

$$\bar{E}_t = \beta_0 + \beta_1 P_t + \beta_z \mathbf{z}_t$$

where $\bar{E}_t$ is the long-run desired energy consumption in period t, P_t is the price of energy in period t, $\mathbf{z}_t$ is the vector of input prices and the output price in period t, and β_0, β_1, and β_z are constant parameters of the demand function.

The long-run demand function shows the amount of energy firms would wish to consume if prices P_t and $\mathbf{z}_t$ were given. However, firms are not usually able to adjust their energy consumption to the desired long-run level *instantaneously*. Instead, it is typically more efficient to adjust gradually to the desired long-run level. We therefore assume that *actual* energy demand adjusts as follows:

$$E_t - E_{t-1} = \alpha(\bar{E}_t - E_{t-1}), \qquad 0 < \alpha < 1$$

This says that the actual adjustment in energy consumption from $t-1$ to t is a fraction, α, of the gap between long-run desired consumption in t and actual consumption in $t-1$. Solving the adjustment equation for $\bar{E}_t$ and substituting it into the energy demand equation gives

$$E_t = (1-\alpha)E_{t-1} + \alpha\beta_0 + \alpha\beta_1 P_t + \alpha\beta_z \mathbf{z}_t \tag{S18.2}$$

In recent years economists have used market data on energy demand, E_t, and prices P_t and $\mathbf{z}_t$ with this model or variants of it to estimate the values of the parameters of this demand equation (α, β_0, β_1, and β_z). Their interest is in determining the response of energy demand to price changes both in the short run and long run. We can use our knowledge of difference equations to work out the dynamics of energy demand adjustment.

Assume that the price of energy changes to a new level and, to simplify notation, assume that this price change occurs in time period 0. The new price level is P_0. Assume further that it and all other prices remain constant thereafter. We now have an autonomous, linear, first-order difference equation for energy demand

$$E_t = (1-\alpha)E_{t-1} + \alpha\beta_0 + \alpha\beta_1 P_0 + \alpha\beta_z \mathbf{z}$$

The *immediate* response of energy demand to the price change is given by the partial derivative

$$\frac{\partial E_0}{\partial P_0} = \alpha\beta_1$$

The *long-run* (steady-state) response of energy demand to the price change is determined by first obtaining an expression for the steady-state level of energy demand. This is found by setting $E_t = E_{t-1} = \bar{E}$. Doing this gives

$$\bar{E} = \beta_0 + \beta_1 P_0 + \beta_z \mathbf{z}$$

Thus the long-run (steady-state) response of energy demand to the price change that occurs in period 0 is given by the partial derivative

$$\frac{\partial \bar{E}}{\partial P_0} = \beta_1$$

As a result the long-run demand response is $1/\alpha$ times larger than the short-run response. For example, if $\alpha = 0.5$, then the long-run response to the price change is twice as large as the short-run response.

To investigate the dynamics of energy demand adjustment more thoroughly, the difference equation can be solved. By theorem 18.1, the solution is

$$E_t = (1-\alpha)^t E_0 + \alpha(\beta_0 + \beta_1 P_0 + \beta_z \mathbf{z})\left(\frac{1-(1-\alpha)^t}{1-(1-\alpha)}\right)$$

where E_0 is energy consumption in time period 0. Simplifying, rearranging, and using the expression for $\bar{E}$ gives

$$E_t = (1-\alpha)^t (E_0 - \bar{E}) + \bar{E}$$

We see clearly now the reason for the restriction $0 < \alpha < 1$ in the original specification of the adjustment equation: only then does energy demand converge to the desired level in the long run. Moreover, since this restriction implies that $0 < 1 - \alpha < 1$, energy demand converges *monotonically* to $\bar{E}$.

Figure S18.1 depicts the adjustment of energy demand toward its steady-state level. At time period 0 we suppose that the price of energy rises to a new level, P_0, causing an immediate reduction in energy demand to E_0 (from its previous level, which we denote $\bar{E}_{-1}$ to represent an implicit assumption that energy demand

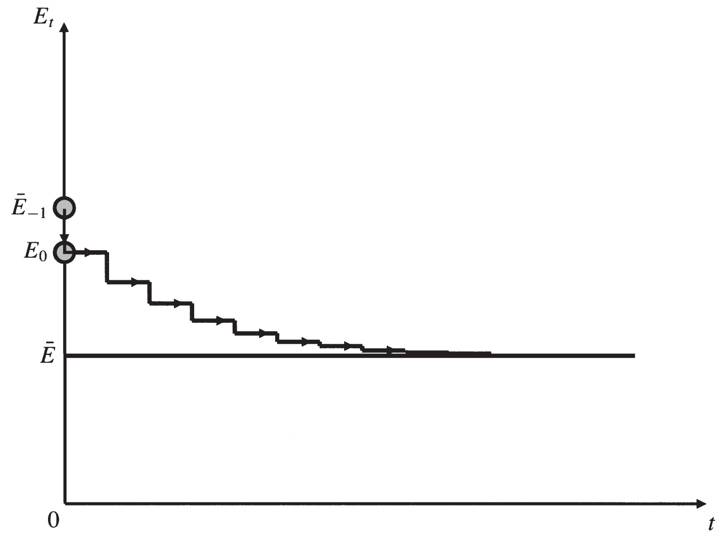

Figure S18.1 Path of energy demand in a partial adjustment model was in a different long-run equilibrium to begin with). Thereafter, energy demand gradually adjusts (converges) toward the desired long-run level.

PRACTICE EXERCISES

S18.1. Solve the difference equation for the modified cobweb model for the parameter values given in exercise 6 from the textbook, assuming that $\theta = 0.6$.

Solutions

S18.1. **(a)** $P_t = -3(-0.2)^t + 5$

(b) $P_t = -1(-0.8)^t + 3$

(c) $P_t = -2(-0.4)^t + 4$

Chapter S19 Nonlinear, First-Order Difference Equations

Contents

S19.1 An Economic Growth Model

Let aggregate output in a model economy be given by

$$y_t = k_t^{\alpha}, \qquad 0 < \alpha < 1;\ t = 0, 1, 2, \ldots$$

where y_t is aggregate output and k_t is the aggregate capital stock. This expression says that the economy's output is a concave function of its stock of productive capital.

Capital is accumulated in this economy by saving (i.e., not consuming) some of the current output. We assume that a constant share, s, of output is saved each period. Assuming further that capital depreciates at the rate δ, we get

$$k_{t+1} = k_t - \delta k_t + s y_t$$

which says that the capital stock in period $t + 1$ is equal to its amount in period t, less depreciation during period t, plus savings from period t. Making the substitution for aggregate output gives a first-order, nonlinear difference equation for the economy's capital stock

$$k_{t+1} = k_t(1 - \delta) + s k_t^{\alpha}, \qquad t = 0, 1, 2, \ldots \tag{S19.1}$$

Given the assumed saving behavior in this model economy, we would like to know as much as possible about the path of the economy's capital stock over time. Does it grow forever? Does it converge to a steady-state value? Does it oscillate?

To answer these questions, we conduct a qualitative analysis of the difference equation, beginning with the phase diagram. For this purpose we need to graph equation (S19.1). To do so, note that $k_{t+1} = 0$ when $k_t = 0$, so the graph goes through the point (0, 0). Next, take the first derivative:

$$\frac{dk_{t+1}}{dk_t} = 1 - \delta + s\alpha k_t^{\alpha-1}$$

which exceeds 0 since we will assume that $\delta < 1$. This result tells us that the curve is upward-sloping for all $k_t > 0$. Next, take the second derivative:

$$\frac{d^2k_{t+1}}{dk_t^2} = s\alpha(\alpha - 1)k_t^{\alpha-2}$$

which is less than 0 because $0 < \alpha < 1$ by assumption. This result, together with the first derivative, tells us that the curve is strictly concave for all $k_t > 0$. This is enough essential information to draw a rough sketch of the curve. Figure S19.1 depicts a curve with these characteristics.

The steady-state equilibrium values are found by setting $k_{t+1} = k_t$. After some simplification, we get

$$\bar{k}[\delta - s\bar{k}^{\alpha-1}] = 0.$$

The solutions are

$$\bar{k} = 0$$

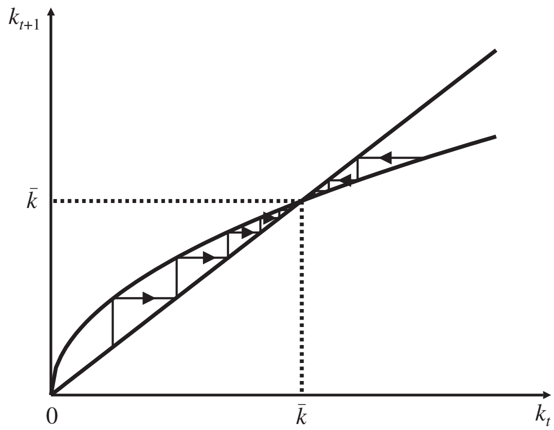

Figure S19.1 Phase diagram for equation (S19.1) showing that $\bar{k}$ is a stable steady state

and

$$\bar{k}^{\alpha-1} = \frac{\delta}{s}$$

Theorem 19.1 requires that we evaluate the derivative of the right-hand side of equation (S19.1) at the steady-state values to determine which of them is stable. This derivative is

$$1 - \delta + s\alpha\bar{k}^{\alpha-1}$$

At $\bar{k} = 0$, this derivative goes to infinity. Hence k_t does not converge to 0 locally. Evaluating the derivative at the second steady state gives

$$1 - \delta + s\alpha\frac{\delta}{s} = 1 - \delta(1 - \alpha)$$

which is larger than 0 and smaller than 1 because both δ and α are between 0 and 1. Hence the point

$$\bar{k} = \left(\frac{\delta}{s}\right)^{1/(\alpha-1)}$$

is locally stable. The phase diagram analysis shows that as long as $k_0 > 0$, then k_t converges to $\bar{k}$. Hence, $\bar{k}$ is globally stable, at least over the positive domain of k. Finally, the analysis we have done shows that the derivative of f is positive over the entire domain of k. Theorem 19.2 tells us then that the approach paths to the steady-state are monotonic, not oscillatory.

S19.2 A Malthusian Growth Model

Thomas Malthus hypothesized that population growth is an inverse function of income per capita up to a biologically determined limit. Let us assume that

$$\frac{N_{t+1} - N_t}{N_t} = n - \frac{b}{w_t}, \qquad t = 0, 1, 2, \ldots \tag{S19.2}$$

where w_t is income per capita, N_t is population in period t, and n and b are positive constants. This says that the rate of population growth per period is equal to n minus a constant divided by income per capita. As w_t rises, the greater food supply and better living conditions mean that the population growth rate rises (because of higher birth rates and lower death rates). The upper limit on the growth rate is n.

Conversely, as w_t falls, the smaller food supply and poorer living conditions mean that the population growth rate declines.

We assume that w_t is given by

$$w_t = \frac{Y_t}{N_t}$$

where Y_t is aggregate output in the economy. For concreteness, we assume that the aggregate production function is

$$Y_t = N_t^{\alpha}, \qquad 0 < \alpha < 1$$

which says that aggregate output is an increasing function of population (labor). After making the appropriate substitution, income per capita is

$$w_t = N_t^{\alpha-1}$$

Substituting this into equation (S19.2) and rearranging gives the nonlinear, first-order difference equation for population in this model

$$N_{t+1} = N_t\left(1 + n - bN_t^{1-\alpha}\right) \tag{S19.3}$$

To construct the phase diagram, we plot equation (S19.3). The function starts at the origin and then increases at a decreasing rate until reaching a maximum; it then decreases at a decreasing rate until eventually intersecting with the horizontal axis. This qualitative information about the phase diagram is found in the usual way as follows: the first derivative is

$$\frac{dN_{t+1}}{dN_t} = 1 + n - b(2-\alpha)N_t^{1-\alpha} \tag{S19.4}$$

which is positive (and equal to $1+n$) when $N_t = 0$ and then decreases monotonically as N_t increases; it reaches zero at N' where

$$N' = \left(\frac{1+n}{b(2-\alpha)}\right)^{1/(1-\alpha)}$$

and then becomes negative. This indicates that the function reaches a maximum at N'; this conclusion is verified by evaluating the second derivative of the function

$$\frac{d^2N_{t+1}}{dN_t^2} = -b(2-\alpha)(1-\alpha)N_t^{-\alpha} < 0$$

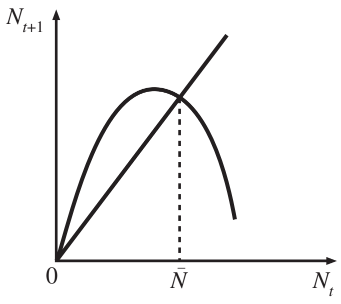

Figure S19.2 Phase diagram for the Malthusian growth model

Figure S19.2 shows the phase diagram for the case in which the steady-state value of N occurs to the right of the peak. Because the phase curve is hill-shaped, the path of N_t in this model could converge to a steady-state point, converge to a stable limit cycle, or could be chaotic. To determine the behavior of the path of N_t in the neighborhood of $\bar{N}$, evaluate the derivative in equation (S19.4) at this point. To do this, we require an analytical solution for $\bar{N}$. Setting $\bar{N} = N_{t+1} = N_t$ in equation (S19.3) and solving gives

$$\bar{N}^{1-\alpha} = \frac{n}{b}$$

Substituting this value into equation (S19.4) and simplifying gives

$$\frac{dN_{t+1}}{dN_t} = 1 - n(1-\alpha)$$

Since $n(1-\alpha) > 0$, the slope is always less than 1. For stability, we also require the slope to be greater than -1, which occurs only if

$$n(1-\alpha) < 2$$

Is this condition likely to be satisfied? We know that $0 < \alpha < 1$. If α were as large as 0.99, n would have to be smaller than 200 to satisfy this condition. This seems likely, as n is the upper limit on the population growth rate per period. If a period is 25 years, say, (about the number of years for one generation), a growth rate of 200 would mean a 200-fold increase in the population every generation, which seems absurdly high. For high values of α, then, we would expect the population to converge to the steady state. If α were as small as 0.01, on the other hand, n would have to be smaller than 2.02 to satisfy the stability condition. This event seems less likely. Indeed, a doubling of the population every generation seems quite possible. Thus, for small values of α, it is less likely that the population will converge to a steady state; instead, it could display cyclical behavior or even chaotic behavior.

PRACTICE EXERCISES

S19.1. For the growth model solve for k_t when $t = 1, 2, 3, 4$ given the following parameter values: $k_0 = 95000$, $s = 0.2$, $\delta = 0.02$, and $\alpha = 0.8$. Find the steady-state value of the capital stock and determine whether k_t converges to it monotonically or in an oscillatory fashion.

S19.2. For the growth model, instead of savings being a constant share of output, assume savings each period are equal to $y_t^{1/2}$. Derive the difference equation for the capital stock, sketch its phase diagram, and determine whether the steady state is stable.

S19.3. Make the following change to the growth model. Instead of aggregate output being a concave function of capital, assume that it is the following linear function:

$$y_t = a + bk_t, \qquad a, b > 0$$

Derive the difference equation for the capital stock. What parameter restrictions are required to ensure the steady-state capital stock is positive? Sketch the phase diagram, and determine whether the steady state is stable.

Solutions

S19.1. $\bar{k} = 100{,}000$, $k_1 = 95{,}019.59$, $k_2 = 95{,}039.11$, $k_3 = 95{,}058.55$, $k_4 = 95{,}077.92$. Since

$$\frac{dk_{t+1}}{dk_t} = 1 - \delta + \frac{s\alpha}{k_t^{1-\alpha}} > 0$$

for all $k > 0$ and $0 < \delta < 1$, then by theorem 19.2, the path of y_t is monotonic. Further, since

$$\left.\frac{dk_{t+1}}{dk_t}\right|_{\bar{y}} = 1 - 0.02 + \frac{(0.2)(0.8)}{(100{,}000)^{0.2}} = 0.996$$

the path of k_t converges to $\bar{k}$.

S19.2. The difference equation becomes

$$k_{t+1} = k_t(1 - \delta) + k_t^{\alpha/2}$$

The steady-state points are the solution to

$$\bar{k}\left[1 - (1 - \delta) - \bar{k}^{\frac{\alpha-2}{2}}\right] = 0$$

which gives $\bar{k} = 0$ and $\bar{k} = \delta^{2/(\alpha-2)}$. Using theorem 19.1, we have

$$\frac{dk_{t+1}}{dk_t} = 1 - \delta + \frac{\alpha}{2}\bar{k}^{\frac{\alpha-2}{2}}$$

Since $0 < \alpha < 1$, this derivative tends toward infinity as $\bar{k} \longrightarrow 0$. So $\bar{k} = 0$ is not stable. At $\bar{k} = \delta^{2/(\alpha-2)}$, the derivative equals $1 - \delta(1 - \alpha/2)$ which is between 0 and 1, so this point is stable. See figure S19.3.

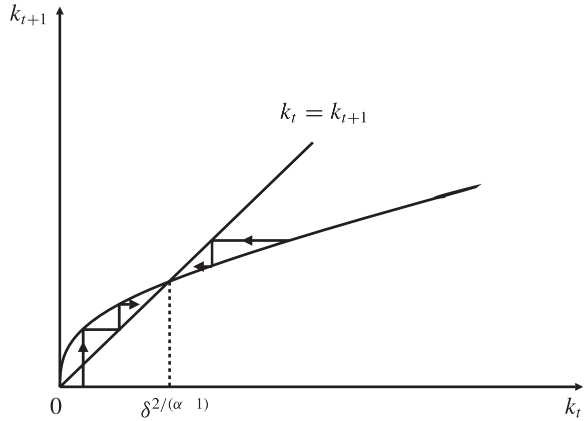

Figure S19.3

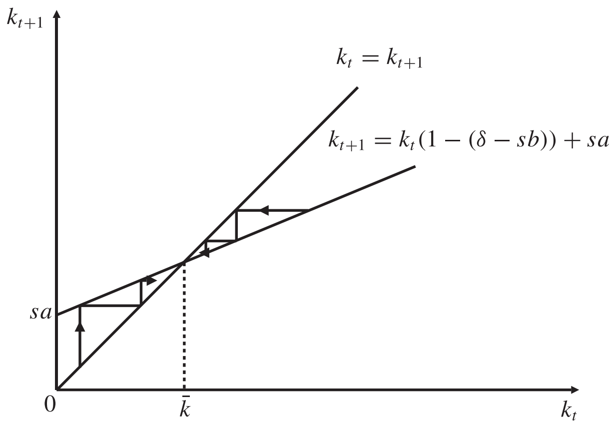

Figure S19.4

S19.3. The difference equation for capital is

$$k_{t+1} = k_t(1-\delta) + sy_t$$

Substituting for y_t and rearranging gives

$$k_{t+1} = k_t(1-\delta+sb) + sa$$

The steady-state is $\bar{k} = sa(\delta - sb)$. We therefore require $(\delta - sb) > 0$ to ensure that $\bar{k} > 0$. Since this is a linear difference equation, we know that k_t converges if and only if the absolute value of $1-\delta+sb$ is less than 1. That is, we require that $-1 < 1-(\delta-sb) < 1$. The restriction that $\delta - sb > 0$ ensures that the upper restriction is met. To ensure that the lower restriction is met, we require that $\delta - sb < 2$. The restriction $\delta - sb > 0$ ensures that the slope of the k_{t+1} line is less than 1 so that it intersects the 45° line. See figure S19.4.

Chapter S20 Linear, Second-Order Difference Equations

Contents

S20.1 A Multiplier-Accelerator Model

Let aggregate national income be given by

$$Y_t = C_t + I_t + G_t$$

where C, I, and G are consumption, investment, and government expenditure, respectively. Assume that G is constant over time at $G_t = \bar{G}$, and assume that consumption is always given by

$$C_t = mY_t, \qquad 0 < m < 1$$

where m is the marginal propensity to consume. Now assume that there is an endogenous component to investment, I^n, and an exogenous component, I^x. For the endogenous component, assume that investors in the economy base their investment decisions for period t on the amount by which national income grew in the previous period. In particular, assume that

$$I_t^n = \alpha(Y_{t-1} - Y_{t-2})$$

where α is a positive constant. Assume that the exogenous component is

$$I_t^x = (1 + g)^t$$

where $g > 0$ is the growth rate of the exogenous part of investment. The total investment is

$$I_t = I_t^n + I_t^x$$

Substitute the investment expression and the consumption expression into the national income identity to get

$$Y_t = mY_t + \alpha(Y_{t-1} - Y_{t-2}) + (1+g)^t + \bar{G}$$

Simplify to obtain

$$Y_t - \frac{\alpha}{1-m}Y_{t-1} + \frac{\alpha}{1-m}Y_{t-2} = \frac{(1+g)^{t+2}}{1-m} + \frac{\bar{G}}{1-m}$$

Write this in more common form by adding 2 to all the time subscripts

$$Y_{t+2} - \frac{\alpha}{1-m}Y_{t+1} + \frac{\alpha}{1-m}Y_t = \frac{(1+g)^t}{1-m} + \frac{\bar{G}}{1-m}$$

To solve this difference equation, we first solve its homogeneous form. This is given by

$$Y_{t+2} - \frac{\alpha}{1-m}Y_{t+1} + \frac{\alpha}{1-m}Y_t = 0$$

The roots of this difference equation are

$$r_1, r_2 = \frac{\alpha}{2(1-m)} \pm \frac{1}{2}\sqrt{\left(\frac{\alpha}{1-m}\right)^2 - \frac{4\alpha}{1-m}}$$

and the solution to the homogeneous version is

$$Y_t = C_1 r_1^t + C_2 r_2^t$$

To find a particular solution, we use the method of undetermined coefficients. Noting that the term in the complete difference equation is

$$\frac{\bar{G}}{1-m} + \frac{(1+g)^{t+2}}{1-m}$$

we guess that the particular solution is

$$y_p = A_0 + A_1(1+g)^{t+2}$$

where A_0 and A_1 are the coefficients to be determined. Since this particular solution must satisfy the complete difference equation, substitute it in to get

$$A_0 + A_1(1+g)^{t+4} - \frac{\alpha}{1-m}[A_0 + A_1(1+g)^{t+3}] + \frac{\alpha}{1-m}[A_0 + A_1(1+g)^{t+2}]$$
$$= \frac{\bar{G}}{1-m} + \frac{(1+g)^{t+2}}{1-m}$$

Factoring out $(1+g)^{t+2}$ on the left-hand side and cancelling terms reduces this to

$$(1+g)^{t+2}\left[(1+g)^2 A_1 - \frac{\alpha(1+g)A_1}{1-m} + \frac{\alpha A_1}{1-m}\right] + A_0 = \frac{\bar{G}}{1-m} + \frac{(1+g)^{t+2}}{1-m}$$

This equality determines the values of the coefficients. For the equality to be true for all values of t, we must set

$$A_0 = \frac{\bar{G}}{1-m}$$

and

$$\left[(1+g)^2 - \frac{\alpha(1+g)}{1-m} + \frac{\alpha}{1-m}\right]A_1 = \frac{1}{1-m}$$

which gives

$$A_1 = \frac{1}{(1+g)^2(1-m) - \alpha g}$$

The general solution to the complete difference equation then is

$$Y_t = C_1 r_1^t + C_2 r_2^t + \frac{\bar{G}}{1-m} + \frac{(1+g)^{t+2}}{(1+g)^2(1-m) - \alpha g}$$

Interestingly it is still possible for Y_t to converge to y_p, even though y_p is not a steady-state solution. If the absolute values of both roots are less than 1, y_t will converge to the particular solution y_p as t goes to infinity. The economic interpretation of this behavior is that, given enough time, national income will converge to a long-run growth path that is given by y_p.

Are the absolute values of the roots less than 1? We will use theorem 20.6 to answer. The first necessary condition is $1 + a_1 + a_2 > 0$. In this model

$$1 + a_1 + a_2 = 1 - \frac{\alpha}{1-m} + \frac{\alpha}{1-m} = 1$$

This condition is clearly satisfied. The second condition is $1 - a_1 + a_2 > 0$. In this model

$$1 - a_1 + a_2 = 1 + 2\frac{\alpha}{1-m}$$

This condition is also satisfied, since $\alpha/(1-m) > 0$ by assumption. The final condition is $a_2 < 1$ which in this model is

$$\frac{\alpha}{1-m} < 1$$

This condition is not satisfied, in general, since the only restrictions placed on α and m in the model are that they each are between 0 and 1. Therefore the model does not, in general, converge to the long-run growth path, unless we impose the additional condition that $\alpha/(1-m) < 1$.

PRACTICE EXERCISES

S20.1. Consider the following multiplier-accelerator model. Assume that consumption is given by

$$C_t = mY_t$$

investment is given by

$$I_t = \alpha(Y_{t-1} - Y_{t-2})$$

and government expenditure grows over time according to

$$G_t = G_0(1+g)^t$$

where g is the exogenous growth rate of government expenditure. Derive and solve the linear, second-order difference equation implied by this model. Show that if the absolute values of the real parts of the characteristic roots are less than 1, then national income converges to an exponential growth path.

S20.2. Consider yet another modification of the multiplier-accelerator model. Assume that consumption is given by

$$C_t = mY_t$$

investment is given by

$$I_t = \bar{I} + \alpha(Y_{t-1} - Y_{t-2})$$

and government expenditure declines over time according to

$$G_t = G_0(1-\delta)^t, \qquad 0 < \delta < 1$$

where δ is the exogenous rate of decline of government expenditure. Derive and solve the linear, second-order difference equation implied by this model. Show that if the absolute values of the real parts of the characteristic roots are less than 1, then national income converges to $\bar{I}/(1-m)$.

S20.3. Suppose that two firms share a growing market. Assume that the (inverse) demand curve is

$$p(x+y) = 120\alpha^t - (x_t + y_t), \qquad \alpha < 1$$

Assuming that each firm makes a Cournot assumption about the other firm's output and maximizes profit (assume that costs are zero), derive and solve the second-order difference equation for x implied by this model.

Solutions

S20.1. Substitute into the national income identity $Y_t = C_t + I_t + G_t$. After re-arranging, we get

$$Y_{t+2} - \frac{\alpha}{1-m}Y_{t+1} + \frac{\alpha}{1-m}Y_t = \frac{G_0}{1-m}(1+g)^{t+2}$$

The homogeneous solution is the same as the example solved in section 20.2. For the particular solution, we try $y_p = A(1+g)^t$. Substitute this solution into the difference equation to get

$$
\begin{aligned}
&A(1+g)^{t+2} - \frac{\alpha}{1-m}A(1+g)^{t+1} + \frac{\alpha}{1-m}A(1+g)^{t} \\
&= \frac{G_0}{1-m}(1+g)^{t+2}
\end{aligned}
$$

Simplifying and solving gives

$$A = \frac{G_0}{(1-m)-\alpha g}$$

The complete solution then is

$$Y_t = C_1(r_1)^t + C_2(r_2)^t + \frac{G_0}{(1-m)-\alpha g}(1+g)^t$$

where

$$r_1,\ r_2 = \frac{\alpha}{2(1-m)} \pm \frac{1}{2}\sqrt{\left(\frac{\alpha}{1-m}\right)^2 - \frac{4\alpha}{1-m}}$$

If the absolute values of the roots are less than 1, then the first two terms in the solution for Y_t tend toward 0 as t gets very large. Thus Y_t converges in the limit to the third term, which produces exponential growth at the rate g.

S20.2. The difference equation is

$$Y_{t+2} - \frac{\alpha}{1-m}Y_{t+1} + \frac{\alpha}{1-m}Y_t = \frac{\bar{I}}{1-m} + \frac{G_0}{1-m}(1-\delta)^t$$

The homogeneous solution is the same as above. For the particular solution, we try $y_p = A_0 + A_1(1-\delta)^t$. After substituting this into the complete difference equation, this gives $A_0 = \bar{I}/(1-m)$ and $A_1 = G_0/[(1-m)(1-\delta)^2 + \alpha\delta]$. The complete solution is

$$
\begin{aligned}
Y_t = C_1(r_1)^t + C_2(r_2)^t &+ \frac{\bar{I}}{1-m} \\
&+ \frac{G_0}{(1-\delta)^2(1-m)+\alpha\delta}(1-\delta)^t
\end{aligned}
$$

If the absolute values of the roots are less than 1, then because $0 < \delta < 1$, Y_t converges in the limit to $\bar{I}/(1-m)$.

S20.3. The reaction functions are $x_{t+1} = 60\alpha^t - y_t/2$ and $y_{t+1} = 60\alpha^t - x_t/2$. After substituting, the second-order difference equation for x_t becomes

$$x_{t+2} - \frac{1}{4}x_t = 60\alpha^t\left(\alpha - \frac{1}{2}\right)$$

The roots are $r_1,\ r_2 = \pm 1/2$. The particular solution we try is $x_p = A\alpha^t$ as long as $\alpha \neq \pm 1/2$, for then the particular solution would have a term in common with the homogeneous solution. The complete solution is

$$x_t = C_1\left(\frac{1}{2}\right)^t + C_2\left(-\frac{1}{2}\right)^t + \frac{60}{\alpha + \frac{1}{2}}\alpha^t$$

Chapter S21 Linear, First-Order Differential Equations

Contents

S21.1 The Dynamics of National Debt Accumulation

Many countries have run persistent budget deficits in recent years. This has led to a dramatic growth in national debts and a concern that this trend could lead to bankruptcy (which would occur if a nation's debt were to become so large that its interest payments exceeded national income). Is this a necessary consequence of persistent deficit financing? Are countries that run persistent deficits on a path toward bankruptcy? These are questions that can be answered by analyzing the dynamics of debt accumulation and income growth. For our purposes we will take a relatively simplified view of these processes to keep the differential equations simple.

Let $D(t)$ represent the dollar value of the debt at time t, and let $Y(t)$ represent the dollar value of the nation's income, or GNP, at time t. We will abstract from inflation by assuming that all variables are denominated in real dollar terms. We will assume that the deficit (defined as a posititive value equal to expenditures minus revenues) is a constant proportion of national income at any point in time. Since the change in the debt is just the deficit, we have

$$\dot{D} = bY, \qquad b > 0 \tag{S21.1}$$

as the ordinary differential equation that describes the behavior of debt. (Typically the value for b in many countries would fall in the range 0.02 to 0.08 which means that deficits are about 2% to 8% of the size of national income). We further assume that national income grows over time according to the following differential equation:

$$\dot{Y} = gY \tag{S21.2}$$

where g is a positive constant (representing the growth rate of national income). Together, equations (S21.1) and (S21.2) are a model of debt accumulation. To analyze the implications of the model for the long-run ratio of interest payments to national income, we need to solve these equations. We start with equation (S21.3), which we can rewrite as

$$\frac{\dot{Y}}{Y} = g$$

Integrating both sides gives

$$\ln Y(t) + c_2 = gt + c_1$$

which we can rewrite as

$$Y(t) = C_1 e^{gt}$$

where $C_1 = e^{c_1 - c_2}$. Assuming that the initial time is $t_0 = 0$ and that the initial values of income and debt are Y_0 and D_0 respectively, we require $Y(0) = Y_0 = C_1$. Thus the solution to the initial-value problem for equation (S21.2) is

$$Y(t) = Y_0 e^{gt} \tag{S21.3}$$

Substitution of this solution into equation (S21.1) gives

$$\dot{D} = bY_0 e^{gt}$$

Although this is actually nonautonomous, it is in a form that can be solved by direct integration. Integrating both sides gives

$$D(t) = bY_0 \frac{e^{gt}}{g} + C_2$$

Since $D(0) = D_0$, the value of C_2 must be set to $(D_0 - b/gY_0)$. Using this value, we have the solution

$$D(t) = D_0 + \frac{b}{g} Y_0 (e^{gt} - 1)$$

Inspection of the solution indicates that national debt, $D(t)$, grows without limit in this model. However, our real concern is with the country's ability to meet the interest obligations on the debt. We assume a constant interest rate r, and we calculate the ratio of interest payments $[rD(t)]$ to national income $Y(t)$ as

$$\frac{rD(t)}{Y(t)} = r\frac{D_0 + bY_0(e^{gt} - 1)/g}{Y_0 e^{gt}}$$

Defining $z(t) \equiv rD(t)/Y(t)$ as the share of national income absorbed by interest payments on the national debt and simplifying produces

$$z(t) = r\frac{D_0}{Y_0}e^{-gt} + r\frac{b}{g}(1 - e^{-gt}) \tag{S21.4}$$

This expression gives the ratio of interest payments to national income at any point in time in this model. Our main interest is to determine whether this ratio converges to a finite limit less than 1 (interest payments never become as large as national income).

Inspection of equation (S21.4) indicates that $z(t)$, the ratio of interest obligations to income, converges to a finite limit as $t \to \infty$. To see this, take the limits of the two terms on the right-hand side as $t \to \infty$, keeping in mind that e^{-gt} goes to 0 as $t \to \infty$. We obtain

$$\lim_{t\to\infty} z(t) = r\frac{b}{g} \tag{S21.5}$$

Interest payments on the debt converge to a constant proportion of national income equal to rb/g. If $rb/g < 1$, then even if a government forever runs a deficit which is a constant proportion of a growing national income, the *burden* on the economy of the resulting debt converges to a constant share of national income. This would be good news because it would mean the economy would always be able to meet its debt payments and bankruptcy would never occur. On the other hand, if $rb/g > 1$, then the process converges to a finite limit where interest payments exceed national income. In this case the economy would be destined to experience bankruptcy if it continued to run deficits.

What is the intuitive explanation for our finding? Because $\dot{D} = bY$ and $\dot{Y} = gY$, the ratio of the increase in debt to the increase in income, $\dot{D}/\dot{Y}$, is just b/g. Thus, for every dollar increase in national income, debt increases by b/g. Suppose that $b/g = 0.5$. Then for every dollar increase in national income, debt increases by 50 cents. Clearly, income is growing faster than the debt, so the ratio of debt to income will always be less than unity. Then, because interest rates are typically much less than 1, the ratio of interest on debt to income will always be less than unity.

On the other hand, suppose that $b/g = 1.5$. Then every dollar increase in income leads to a \$1.50 increase in debt. Debt is now growing faster than income, so the ratio of debt to income will definitely exceed unity eventually. In this case interest on the debt could also exceed national income if the interest rate is high enough.

Some typical values for the ratio b/g can be calculated from data contained in the *OECD Economic Outlook*. For the United States, the ratio averaged 0.61 over the high-growth years 1987–89, rose above 1 in the 1990s and averaged 0.81 from 2001 to 2006. However, the combination of high military spending and high deficits due to the financial crisis that began in 2008 pushed the ratio up to 2.89 in 2010. Italy had a ratio exceeding 3.0 over the late 1980s and reaching a dangerously high value of 12.0 by 1990. Fortunately, Italy's b/g dropped to an average of 0.9 from 2001 to 2006. The United Kingdom had a very high ratio, 12.0, in 1990, but greater fiscal prudence brought the ratio down to an average of 0.63 from 2001 to 2006. Other countries, such as Australia, Canada, and Korea, ran budgetary surpluses on average from 2001 to 2006 so that their b/g ratios were negative, indicating that national debt was actually falling.

For some of these countries the ratio b/g was substantially larger than 1 in some years, which is cause for concern. Concern would be even higher if these countries had been unable to bring the ratio down to more sustainable levels as was common in the period of strong economic growth from 2001 to 2006. Nevertheless, by analyzing the dynamics of this model, we have discovered the surprising result that even if governments do run persistent deficits, even large ones, bankruptcy is not a necessary long-run consequence. However, different specifications of the model of deficit spending can produce less optimistic results. See exercise S21.3 for an example.

S21.2 The Dynamics of the IS-LM Model

Consider the following IS-LM model:

$$
\begin{aligned}
C &= a + bY - lR && \text{(consumption demand)}\\
I &= \bar{I} && \text{(investment demand)}\\
G &= \bar{G} && \text{(government demand)}\\
L &= kY - hR && \text{(money demand)}\\
M &= \bar{M} && \text{(money supply)}
\end{aligned}
$$

The endogenous variables in this model are output (Y) and the interest rate (R). Equilibrium in the goods market requires that aggregate demand ($C + I + G$) equal aggregate supply (Y). Equilibrium in the money market requires that the demand for money (L) equal the supply of money (M). When these two conditions hold, the equilibrium values of output and interest rate are determined.

We will assume that the money market clears instantly (R adjusts instantaneously to equate the demand for and supply of money). However, we do not assume this about the goods market. Instead, we assume that output adjusts gradually in response to the demand-supply gap. In particular, we assume that

$$\dot{Y} = \alpha(a + bY - lR + \bar{I} + \bar{G} - Y), \qquad \alpha > 0 \tag{S21.6}$$

where α is a positive coefficient determining the speed at which the goods market adjusts. This model gives us a linear, first-order differential equation for Y. We wish to solve this for $Y(t)$ and then determine whether the equilibrium is stable.

Rewrite the differential equation as

$$\dot{Y} - \alpha(b-1)y + \alpha lR = \alpha(a + \bar{I} + \bar{G})$$

Because the money market clears instantly, R will always be at its equilibrium value. Equating the demand for money with the supply of money gives

$$R = \frac{k}{h}Y - \frac{\bar{M}}{h}$$

Making the substitution for R and simplifying gives

$$\dot{Y} + \alpha\left(1 - b + \frac{lk}{h}\right)Y = \alpha\left(a + \bar{I} + \bar{G} + \frac{l\bar{M}}{h}\right)$$

To simplify the next few steps, define A and B as

$$A = \alpha\left(1 - b + \frac{lk}{h}\right)$$

$$B = \alpha\left(a + \bar{I} + \bar{G} + \frac{l\bar{M}}{h}\right)$$

The differential equation then becomes

$$\dot{Y} + AY = B$$

Assuming that the initial condition for output is $Y(0) = Y_0$, then applying equation (21.8) from chapter 21 of the textbook gives the complete solution

$$Y(t) = \left(Y_0 - \frac{B}{A}\right)e^{-At} + \frac{B}{A}$$

As usual, the solution shows that if Y_0 happens to be equal to its steady-state value, it will always equal the steady-state value.

What happens if Y_0 is not equal to the steady-state value of Y? The solution shows that $Y(t)$ converges to its steady state, B/A, if and only if $A > 0$. We conclude that the IS-LM model contains a stable equilibrium if and only if $A > 0$. By the definition of A, this requires that

$$1 - b + \frac{lk}{h} > 0$$

Provided that $l > 0$, which would normally be assumed, we can rewrite this as

$$\frac{k}{h} > -\frac{1-b}{l}$$

In this form the stability condition requires that the slope of the LM curve (the left-hand side) exceed the slope of the IS curve (the right-hand side). This condition is met if the parameters satisfy the usual conditions ($0 < b < 1, l, k, h > 0$) which produce a positively sloped LM curve and a negatively sloped IS curve. However, even if these usual conditions are not satisfied, the equilibrium could still be satisfied. For example, the LM curve can be negatively sloped ($k/h < 0$) provided that it is flatter than the IS curve.

PRACTICE EXERCISES

S21.1. Using the IS-LM model, assume the goods market clears instantly, making aggregate demand always equal to aggregate supply. However, assume that the interest rate adjusts gradually at speed α in response to the gap between the demand for money and the supply of money. Derive the differential equation for R. Solve for $R(t)$, and determine the condition on the parameters that must be satisfied for the equilibrium to be stable.

S21.2. Modify the IS-LM model by letting investment demand depend on the interest rate. That is, assume that

$$I = \bar{I}(1 - R)$$

and assume that $0 < R < 1$. Derive the differential equation for Y. Solve for $Y(t)$ and determine the condition the parameters must satisfy for the equilibrium to be stable.

S21.3. Suppose that the government runs a deficit, gross of interest payments on the debt, which is a fixed proportion b of national income. Then debt increases by the amount of the deficit plus the interest payments on the debt. If r denotes the constant interest rate paid on the debt, then the differential equation for debt is

$$\dot{D} = bY + rD$$

Assuming that $\dot{Y} = gY$, $D(0) = D_0$, and $Y(0) = Y_0$, solve for the ratio $rD(t)/Y(t)$. Show that this ratio converges to a finite limit if and only if the growth rate of income exceeds the interest rate.

S21.4. Suppose that a government always runs a budgetary deficit equal to 12% of national income including interest payments. If national income is growing at a rate of 3%, will this government always be able to meet its interest payments on the debt? If the interest rate is a constant 10%, what share of national income will go toward servicing (paying interest on) the national debt in the limit?

Solutions

S21.1. Instantaneous clearing in the goods market ensures that $C + \bar{I} + \bar{G} = Y$. This gives equilibrium Y as

$$Y = \frac{a - lR + \bar{I} + \bar{G}}{1 - b}$$

The interest rate adjusts according to $\dot{R} = \alpha(L - M)$, which gives

$$\dot{R} = \alpha(kY - hR - \bar{M})$$

Substituting and re-arranging gives

$$\dot{R} + \alpha\left(h + \frac{kl}{1 - b}\right)R = \frac{\alpha k}{1 - b}(a + \bar{I} + \bar{G}) - \alpha\bar{M}$$

The solution is

$$R(t) = (R_0 - \bar{R})e^{-At} + \bar{R}$$

where

$$\bar{R} = \frac{k(a + \bar{I} + \bar{G}) - (1 - b)\bar{M}}{h(1 - b) + kl}$$

and

$$A = \alpha\left(h + \frac{kl}{1-b}\right).$$

Convergence requires that $A > 0$.

S21.2. The money market clears instantly and the equilibrium interest rate is $R = kY/h - \bar{M}/h$. The goods market adjusts according to

$$\dot{Y} = \alpha[a + bY - lR + \bar{I}(1-R) + \bar{G} - Y]$$

Substituting for R and simplifying gives

$$\dot{Y} + AY = B$$

where

$$A = \alpha\left[1 - b + \frac{(l+\bar{I})k}{h}\right]$$
$$B = \alpha\left[a + \frac{(l+\bar{I})\bar{M}}{h} + \bar{I} + \bar{G}\right]$$

We require that $A > 0$ for stability. The solution is

$$Y(t) = \left(Y_0 - \frac{B}{A}\right)e^{-At} + \frac{B}{A}$$

S21.3. Since $\dot{Y} = gY$, we have $Y(t) = Y_0 e^{gt}$. Substituting this into the differential equation for debt gives

$$\dot{D} = bY_0 e^{gt} + rD$$

Re-arranging gives

$$\dot{D} - rD = bY_0 e^{gt}$$

The integrating factor is e^{-rt}. Multiply both sides by this factor and rewrite the equation as

$$\frac{d}{dt}(De^{-rt}) = bY_0 e^{(g-r)t}$$

Integrating gives

$$De^{-rt} = \frac{bY_0 e^{(g-r)t}}{g-r} + C$$

which, after re-arranging, becomes

$$D(t) = \frac{bY_0 e^{gt}}{g-r} + Ce^{rt}$$

At $t = 0$, $D(0) = D_0$. This gives $C = D_0 - bY_0/(g-r)$. The solution for debt becomes

$$D(t) = \frac{bY_0 e^{gt}}{g-r} + \left(D_0 - \frac{bY_0}{g-r}\right)e^{rt}$$

The ratio of interest payments, $rD(t)$, to national income is

$$\frac{rD(t)}{Y(t)} = \frac{rb}{g-r} + e^{(r-g)t}\left(\frac{rD_0}{Y_0} - \frac{rb}{g-r}\right)$$

If $g > r$, the exponential term goes to 0 in the limit and the solution converges to the finite limit $rb/(g-r)$. If $g < r$, the ratio of interest payments to national income grows without limit. To see this, it helps to rewrite our expression as

$$\frac{rD(t)}{Y(t)} = \frac{rb}{r-g}\left(e^{(r-g)t} - 1\right) + e^{(r-g)t}\frac{rD_0}{Y_0}$$

With $r - g > 0$, both terms on the right-hand side are positive and grow without limit.

S21.4. With $\dot{D} = 0.12Y$ and $\dot{Y} = 0.03Y$, we get

$$\frac{rD(t)}{Y(t)} = \frac{rD_0}{Y_0}e^{-0.03t} + \frac{r(0.12)}{0.03}(1 - e^{-0.03t})$$

If $r = 0.10$, then in the limit as $t \longrightarrow \infty$, $rD/Y = 0.4$. That is, 40% of national income will be used to service the debt.

Chapter S22 Nonlinear, First-Order Differential Equations

Contents

S22.1 Nonlinear Differential Equation Examples

Example S22.1 **A Fishery Model with the Harvest Rate Proportional to Stock Size**

Consider a fish species that is caught with nets. The quantity of fish caught is assumed to be a fraction, α, of the fish population. The size of α will depend on the number of fishing boats used to deploy the nets, but here we will take α as exogenously given. If the growth function of the fishery in the absence of any fishing is

$$g(y) = y - y^2, \quad y(0) = y_0 > 0$$

solve for the fish population as a function of t when the harvest rate is

$$h = \alpha y$$

Solution

The differential equation governing the fish population now becomes

$$\dot{y} = y - y^2 - \alpha y \tag{S22.1}$$

To solve, first we rewrite the equation as

$$\dot{y} - y(1 - \alpha) = -y^2$$

Multiply through by y^{-2} to obtain

$$y^{-2}\dot{y} - y^{-1}(1 - \alpha) = -1$$

Now define $x = y^{-1}$ (note that we require $y(t) \neq 0$ here) which transforms our differential equation into

$$-\dot{x} - x(1-\alpha) = -1$$

or, after multiplying through by -1,

$$\dot{x} + x(1-\alpha) = 1$$

This is a linear, first-order differential equation with a constant coefficient, $1-\alpha$, and a constant term, 1. To solve, rewrite its homogeneous form as

$$\frac{\dot{x}}{x} = -(1-\alpha)$$

The solution to the homogeneous form is

$$x_h = Ce^{-(1-\alpha)t}$$

The particular solution, given by the steady state, is

$$\bar{x} = \frac{1}{1-\alpha}$$

The complete solution is then

$$x(t) = Ce^{-(1-\alpha)t} + \frac{1}{1-\alpha}$$

Using the initial condition $x(0) = x_0$ to solve for C gives

$$x(t) = \left(x_0 - \frac{1}{1-\alpha}\right)e^{-(1-\alpha)t} + \frac{1}{1-\alpha}$$

Now substitute y^{-1} for x and simplify to get

$$y(t) = \frac{y_0(1-\alpha)}{y_0 + (1-\alpha-y_0)e^{-(1-\alpha)t}}$$

This is the solution we are looking for. It shows the size of the population as a function of t. We can go one step further to determine the steady-state value of the fish stock by taking the limit of our solution as $t \to \infty$. We see that $y(t) \to (1-\alpha)$

as $t \to \infty$. We conclude that if the fishing technology is such that the harvest is a constant fraction, α, of the fish population, the size of the fish population will approach a fraction, $1-\alpha$, of its natural equilibrium size, where its natural equilibrium size is 1 as example S22.2 shows. ■

Example S22.2 An Explicit Solution of the Fishery Model

Solve differential equation (22.3) explicitly.

Solution

Note that equation (22.3) is the same as equation (S22.1) if we set $\alpha = 0$. Since we have already obtained a solution for equation (S22.1), we merely need to set $\alpha = 0$ in the solution to obtain the solution for equation (22.3). Doing so gives

$$y(t) = \frac{y_0}{y_0 + (1-y_0)e^{-t}}$$

It is apparent that the limit of $y(t)$ as $t \to \infty$ is 1. This confirms the results of the qualitative analysis. Note that we rule out the other steady-state value of $y = 0$ when we use the technique shown here to obtain an explicit solution to the Bernoulli equation because we cannot allow $y(t) = 0$, given that the technique involves a transformation of variables in which we divide by $y(t)$. ■

The advantage of the explicit solution over the qualitative analysis is that we can use it to calculate the actual value of $y(t)$ at any time t, whereas the qualitative solution does not provide that kind of information. If we need that kind of quantitative information and we cannot obtain an explicit solution, it would be necessary to do a numerical approximation. There are a number of computer software programs now available for personal computers that are designed to do just that.

Example S22.3 The Aggregate Growth Model with Technological Change and Zero Population Growth

In this example we consider a special version of the Solow growth model examined in chapter 22 of the textbook. We assume here that output per person is a function not only of the capital–labor ratio but also of t, which represents the technological improvements that occur over time. In particular, let

$$y = e^{at}k^{1/2}$$

where $a > 0$ is a constant. This says that output is a concave function of the capital–labor ratio, and an increasing function of time, owing to technological progress.

Assuming zero population growth ($n = 0$), the differential equation for k becomes

$$\dot{k} = se^{at}k^{1/2}$$

This is a nonlinear differential equation describing the growth of the capital–labor ratio. We wish to solve this equation to obtain an expression showing k as a function of t. To do this, we notice that the differential equation is separable in t and k because we can write it as

$$-se^{at} + k^{-1/2}\dot{k} = 0$$

where the first term depends only on t and the second term, which multiplies $\dot{k}$, depends only on k.

We re-express this differential equation as

$$-se^{at}\,dt + k^{-1/2}\,dk = 0$$

and integrate directly to obtain

$$-\int se^{at}\,dt + \int k^{-1/2}\,dk = 0$$

Carrying out the integration gives

$$-\frac{s}{a}e^{at} + 2k^{1/2} = C$$

Solving for k gives

$$k(t) = \left(\frac{s}{2a}e^{at} + \frac{C}{2}\right)^2$$

Next we can determine the constant of integration in the usual way from the initial condition on the capital stock, $k(0) = k_0$. This gives

$$C = 2\sqrt{k_0} - \frac{s}{a}$$

The solution becomes

$$k(t) = \left(\frac{s}{2a}(e^{at} - 1) + \sqrt{k_0} \right)^2$$

It is apparent that, in this model, $k(t)$ grows without bound, provided $a > 0$. ■

PRACTICE EXERCISES

S22.1. In the growth model, the differential equation for the economy's capital–labor ratio is

$$\dot{k} = sf(k) - nk$$

where $f(k)$ gives output per worker, y, as a function of the capital–labor ratio, k; s is the saving rate in the economy; and n is the growth rate of the labor force. Let

$$f(k) = k^{1/2}$$

and obtain an explicit solution for k as a function of t. Show that $k(t)$ converges to the steady-state equilibrium point as long as n is positive.

S22.2. In the Solow growth model, allow for technological progress by redefining the labor force as the *effective* labor force, E, which includes not only the number of workers but also the impact of technological improvement. Assuming that the effective labor units per person grow at the rate λ, we have

$$E(t) = L(t)e^{\lambda t}$$

Assuming that the labor force is initially L_0 and grows at the rate n, then we have

$$E(t) = L_0 e^{(n+\lambda)t}$$

The production function becomes $y = f(k)$, where $y = Y/E$ and $k = K/E$. Derive the nonlinear differential equation for k in this augmented model. Find the steady-state equilibrium value of k and determine the convergence property of the model by conducting a qualitative analysis. Finally show that this augmented model implies that output per person, Y/L, grows at the rate λ in the steady state.

Solutions

S22.1. We are given $\dot{k} = sk^{1/2} - nk$. Set $\dot{k} = 0$ to find the steady state. For $k \neq 0$, this gives $\bar{k} = (s/n)^2$. Now write the differential equation as $\dot{k}+nk = sk^{1/2}$ and multiply through by $k^{-1/2}$. This gives $\dot{k}k^{-1/2}+nk^{1/2} = s$. Define $x = k^{1/2}$ so that $\dot{x} = k^{-1/2}\dot{k}/2$. Substituting transforms the differential equation into

$$\dot{x} + \frac{nx}{2} = \frac{s}{2}$$

The solution is $x(t) = Ce^{-nt/2} + s/n$. Since $k = x^2$, this becomes

$$k(t) = \left(Ce^{-nt/2} + \frac{s}{n}\right)^2$$

If $n > 0$, then $k(t)$ converges to $(s/n)^2$ in the limit.

S22.2. Since $k = K/E$ we have $\dot{k} = \dot{K}/E - k\dot{E}/E = sY/E - k(n + \lambda) = sf(k) - (n + \lambda)k$. The steady state occurs at $\dot{k} = 0$, which gives

$$\frac{f(\bar{k})}{\bar{k}} = \frac{n + \lambda}{s}$$

The phase diagram looks very much like that in the Solow growth model in figure 22.4. It indicates that $k(t)$ converges to $\bar{k}$ in the limit. In the steady state, $Y/E = f(k)$ is constant. Therefore $\dot{Y}/E - Y\dot{E}/E^2 = 0$ so that $\dot{Y}/Y = \dot{E}/E = n + \lambda$. But the growth rate of Y/L equals $\dot{Y}/Y - \dot{L}/L = (n + \lambda) - n$. Therefore Y/L grows at the rate λ.

Chapter S23 Linear, Second-Order Differential Equations

Contents

S23.1 A Walrusian Price-Adjustment Model with Entry and Exit

Suppose that price adjusts according to whether there is excess demand or supply at the current price in a competitive market, as in the Walrasian price-adjustment model analyzed in chapter 21. If q^D and q^S are the quantities demanded and supplied respectively, and α is a positive constant, then the price adjusts according to

$$\dot{p} = \alpha(q^D - q^S) \tag{S23.1}$$

In addition suppose that firms enter or exit the industry according to whether economic profits are positive or negative. Let N represent the number of firms in the industry (assume that N is differentiable), and let $\bar{c}$ be a positive constant representing the minimum average cost that firms can achieve. If price exceeds $\bar{c}$, positive economic profits are earned, and this stimulates entry to the industry: $\dot{N} > 0$. If price is less than $\bar{c}$, economic losses are earned, and this stimulates exits from the industry: $\dot{N} < 0$. This is expressed algebraically as follows:

$$\dot{N} = \gamma(p - \bar{c}) \tag{S23.2}$$

where γ is a positive constant that represents the speed at which N adjusts to profits and losses. Let us assume that the demand curve is given by

$$q^D = A + Bp$$

We assume that the demand curve slopes downward in this model, which means we impose the restriction that $B < 0$. We assume that the supply curve in this model is

$$q^S = mN$$

where m is a positive constant. This gives us a supply curve (in a price-quantity diagram) that is vertical for a given N; that is, it is perfectly inelastic with respect to price for a given number of firms.

Our goal is to analyze the path of price over time and, in particular, to determine whether, and if so how, it converges to an equilibrium. Equation (S23.1) is a first-order differential equation for price. However, we cannot use our first-order techniques to solve it because it depends on N, which in turns depends on p. There are two ways to deal with this problem. One way is to treat equations (S23.1) and (S23.2) as a system of two first-order differential equations and solve them simultaneously. We do this in chapter 24. The other way is to differentiate equation (S23.1) with respect to t to get a second-order differential equation. Doing this gives us

$$\ddot{p} = \alpha(B\dot{p} - m\dot{N})$$

Use equation (S23.2) to substitute for $\dot{N}$ to get the following linear, second-order differential equation:

$$\ddot{p} - \alpha B\dot{p} + \alpha m\gamma p = \alpha m\gamma\bar{c}$$

To solve, begin with the homogeneous form

$$\ddot{p} - \alpha B\dot{p} + \alpha m\gamma p = 0$$

The characteristic equation is

$$r^2 - \alpha Br + \alpha m\gamma = 0$$

and

$$r_1,\ r_2 = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_2}}{2}$$

where

$$a_1 = -\alpha B \quad \text{and} \quad a_2 = \alpha m\gamma$$

Assume for now that the roots are distinct. The homogeneous solution is

$$p_h = C_1 e^{r_1 t} + C_2 e^{r_2 t}$$

The particular solution we use is the steady-state price. This is found by setting $\ddot{p} = \dot{p} = 0$. This gives

$$p_p = \bar{c}$$

The steady-state equilibrium value of price is $\bar{c}$, the minimum average cost.

The complete solution is the sum of the homogeneous solution and the particular solution. After using the expressions for a_1 and a_2 and simplifying, the complete solution becomes

$$p(t) = C_1 e^{r_1 t} + C_2 e^{r_2 t} + \bar{c}$$

where

$$r_1,\ r_2 = \frac{\alpha B \pm \sqrt{\alpha^2 B^2 - 4\alpha m \gamma}}{2}$$

The roots in this solution are never positive as long as the demand curve is negatively sloped ($B < 0$) because α, m, and γ are positive constants. If the roots are real-valued, the price path will converge to the steady-state equilibrium, $\bar{c}$. If the roots are complex-valued ($\alpha^2 B^2 - 4\alpha m \gamma < 0$), then theorem 23.4 gives the solution as

$$p(t) = e^{ht}(A_1 \cos vt + A_2 \sin vt) + \bar{c}$$

where

$$h = \frac{\alpha B}{2} \quad \text{and} \quad v = \frac{\sqrt{4\alpha m \gamma - \alpha^2 B^2}}{2}$$

The price path still converges to $\bar{c}$ since the real part of the complex roots, h, is negative, given our assumption that $B < 0$.

Whether the roots are real or complex, price always converges to the steady-state value, $\bar{c}$, in this model. The difference is that the *path* toward this equilibrium is monotonic in the case of real-valued roots, such as the path depicted in figure 23.6 in the textbook, and oscillatory in the case of complex-valued roots, such as the path depicted in figure 23.7.

In the case of complex roots, the price path displays dampened oscillations in its convergence to the equilibrium price. Overshooting of the equilibrium price occurs frequently. This is a fascinating behavior because it indicates that even when price reaches the equilibrium value, it will not necessarily remain there. The reason is that there are two forces operating to change price. Therefore two conditions, not just one, must be satisfied for price to be stationary: quantity supplied must equal quantity demanded and economic profits must be zero.

It is interesting to note that one of the factors more likely to cause the roots to be complex and the price path to display oscillations is a large value of γ. A large value of γ means that the number of firms in the industry adjusts rapidly to realizations of profits or losses. Thus we would expect industries for which entry and exit can occur quickly to be more likely to display fluctuating price behavior than otherwise.

PRACTICE EXERCISES

S23.1. In the *Walrusian price-adjustment model with entry and exit,* find the complete solution for the following parameter values: $\alpha = 0.5, \gamma = 5/4, B = -2, m = 2, \bar{c} = 10$, and $A = 30$.

S23.2. Solve for price as a function of time in the following price-adjustment model. Assume price adjusts to the supply-demand gap according to

$$\dot{p} = \alpha(q^D - q^S)$$

where $q^D = A + Bp$, and with $A > 0$, $B < 0$ and $q^S = Gp + mk$ and with $G > 0, m > 0$. Here, k is the stock of capital (plant and equipment) invested in the industry. More capital means larger supply at the given price. Assume that capital adjusts according to whether firms earn economic profits or losses as follows:

$$\dot{k} = \gamma(p - \bar{c})$$

where $\bar{c} > 0$ is the average cost of production and $\gamma > 0$ is a speed-of-adjustment parameter.

Solutions

S23.1. $r_1,\ r_2 = -1/2 \pm i$; $\bar{p} = 10$.

$$p(t) = e^{-t/2}[A_1 \cos t + A_2 \sin t] + 10$$

S23.2. Differentiating $\dot{p}$ and substituting for $\dot{k}$ gives

$$\ddot{p} - \alpha(B - G)\dot{p} + \alpha m\gamma p = \alpha m\gamma\bar{c}$$

Let

$$
\begin{aligned}
a_1 &= -\alpha(B - G)\\
a_2 &= \alpha m \gamma\\
b &= \alpha m \gamma \bar{c}
\end{aligned}
$$

The solution is

$$p(t) = C_1 e^{r_1 t} + C_2 e^{r_2 t} + \bar{p}$$

where $\bar{p} = b/a_2 = \bar{c}$ and

$$r_1,\ r_2 = \alpha(B - G)/2 \pm \sqrt{\alpha^2(B - G)^2 - 4\alpha m \gamma}/2$$

Chapter S24 Simultaneous Systems of Differential and Difference Equations

Contents

S24.1 A Walrusian Price-Adjustment Model with Entry

In chapter S23 we considered a model of a competitive market in which price adjusts to excess demand or supply and firms enter or exit the industry if profits or losses are being made. We re-examine the model here as a *system* of two linear differential equations rather than as a single second-order differential equation for price.

Price adjusts to excess demand according to

$$\dot{p} = \alpha(q^D - q^S), \qquad \alpha > 0$$

where $q^D = a + bp$ is the demand function, $q^S = mN$ is the supply function, p is price, N is the number of firms in the industry, and α is a speed-of-adjustment coefficient. Making these substitutions gives

$$\dot{p} = \alpha(a + bp - mN) \tag{S24.1}$$

The number of firms adjusts according to

$$\dot{N} = \gamma(p - \bar{c}), \qquad \gamma > 0 \tag{S24.2}$$

where $\bar{c}$ is the fixed average cost of production. Firms enter ($\dot{N} > 0$) if price exceeds average cost (profits positive) and exit if price is less than average cost (profits negative). Together, equations (S24.1) and (S24.2) form the system of two linear, first-order differential equations in this model. We re-express the model as

$$\begin{bmatrix} \dot{p} \\ \dot{N} \end{bmatrix} = \begin{bmatrix} \alpha b & -\alpha m \\ \gamma & 0 \end{bmatrix} \begin{bmatrix} p \\ N \end{bmatrix} + \begin{bmatrix} \alpha a \\ -\gamma \bar{c} \end{bmatrix} \tag{S24.3}$$

Using theorem 24.2 the solutions are

$$p(t) = C_1 e^{r_1 t} + C_2 e^{r_2 t} + \bar{p} \tag{S24.4}$$

$$N(t) = \frac{r_1 - \alpha b}{-\alpha m} C_1 e^{r_1 t} + \frac{r_2 - \alpha b}{-\alpha m} C_2 e^{r_2 t} + \bar{N} \tag{S24.5}$$

where $\bar{p}$ and $\bar{N}$ are the steady-state price and number of firms respectively. The roots of the characteristic equation are

$$r_1, r_2 = \frac{\alpha b}{2} \pm \frac{1}{2}\sqrt{\alpha^2 b^2 - 4\alpha m \gamma}$$

The determinant of the coefficient matrix is

$$\alpha m \gamma > 0$$

The positive determinant indicates that the roots are either both negative or both positive if real valued. The trace of the coefficient matrix is

$$\operatorname{tr}(A) = \alpha b$$

Both roots are negative if and only if $\alpha b < 0$. Since $\alpha > 0$, the only way this can happen is if $b < 0$. Therefore the necessary and sufficient condition for stability in this model is $b < 0$, which is the requirement that demand be negatively sloped. To determine whether the roots are real or complex we calculate $[\operatorname{tr}(A)]^2 - 4|A|$. If this expression is negative, the roots are complex; if it is zero, the roots are real and equal; if it is positive, the roots are real. We get

$$[\operatorname{tr}(A)]^2 - 4|A| = \alpha^2 b^2 - 4\alpha m \gamma$$

Whether this value is positive or negative cannot be determined in general but will depend on the particular numerical values of the parameters α, b, m, and γ. Thus, in general, all we can say about this model is that it converges to the steady state if and only if $b < 0$. The steady state is either a stable node, a stable focus, or an improper stable node, depending on whether $\alpha^2 b^2 - 4\alpha m \gamma$ is positive, negative, or zero. The phase diagram for this model is shown in figure S24.1 for the case

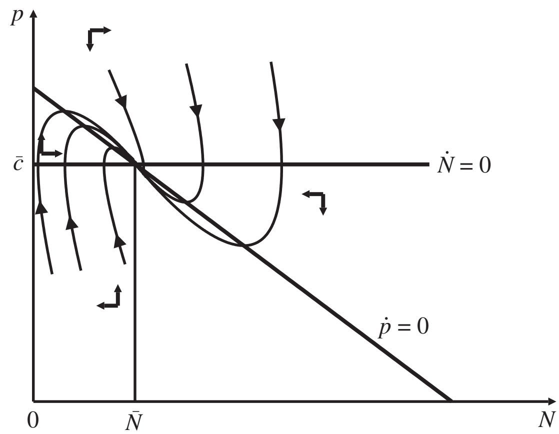

Figure S24.1 Phase diagram for the Walrusian price-adjustment model with entry

of real and distinct (negative) roots, which gives a stable node. To construct this diagram, begin by plotting the isocline for p. Setting $\dot{p} = 0$ and simplifying gives

$$p = \frac{m}{b}N - \frac{a}{b}$$

The$p$ isocline is a straight line with slope m/b, which is negative given $b < 0$, $m > 0$, and intercept $-a/b$, which is positive given $a > 0$.

To determine the motion of p in the two isosectors, use the fact that

$$\frac{\partial \dot{p}}{\partial N} = -\alpha m < 0$$

Therefore moving from any point on the isocline to a smaller (larger)N increases (decreases) $\dot{p}$. Thus $\dot{p} > 0$ to the left of the isocline and $\dot{p} < 0$ to the right of it.

Next plot the isocline for N. Setting $\dot{N} = 0$ and simplifying gives

$$p = \bar{c}$$

which is a line with zero slope at the intercept $\bar{c}$. The motion of N in the two isosectors separated by the N isocline can be determined directly by inspection of the $\dot{N}$ equation. Clearly, $\dot{N} > 0$ for $p > \bar{c}$ and $\dot{N} < 0$ for $p < \bar{c}$.

The phase diagram shows that regardless of where the dynamic system starts, it converges asymptotically to the steady-state equilibrium values of $\bar{p}$ and $\bar{N}$. In the "long run" therefore price converges to average cost.

S24.2 A Markov Model of Layoffs

This model of the transition of laid-off workers between the states of unemployment and re-employment provides a good illustration of how systems of difference equations can arise in economics. Suppose that 1,000 workers at a large manufacturing plant are laid off (because the company is downsizing, say). Only a small proportion of them are able to find another job immediately. Let this proportion be called e_0, where the subscript stands for $t = 0$. The remaining proportion become unemployed and begin searching for another job. Let this proportion be called u_0 (of course, $u_0 = 1 - e_0$).

Let the probability of finding a job during any time period be β. On average, we would expect a proportion β of the unemployed workers to become employed during a time period. However, those workers who were already employed may become unemployed again (their new employer may lay them off or they may quit to find a better job). Let the probability of becoming unemployed during any time period be α. On average we would expect a proportion α of the employed workers to become unemployed during a time period.

The initial layoff of 1,000 workers starts a chain of events. During every time period after the initial layoff, there is a transition of some people from the state of unemployment to the state of employment as well as a transition of some other people from the state of employment to the state of unemployment. Suppose that 100 of the 1,000 laid off are re-employed immediately. Then at $t = 0$, we have 900 unemployed and 100 employed. Suppose that the probability of finding a job is 0.20. Then during $t = 1$, of the 900 unemployed, 180 find jobs and 720 remain unemployed. Suppose further that the probability of losing a job if you have one is 0.05. Then during $t = 1$, of the 100 who had jobs, 5 become unemployed and 95 keep their jobs. We now have 725 unemployed in total and 275 employed in total. During $t = 2$, 20% of the 725 unemployed find jobs and the remaining 145 stay unemployed. Likewise, of the 275 that were employed, 14 become unemployed and the remaining 261 keep their jobs. At the end of period 2 then, we have 596 unemployed and 406 employed. Notice that the proportion of the original 1,000 employees who are employed increases from 0.1 at $t = 0$, to 0.275 at $t = 1$, and to 0.406 at $t = 2$. It turns out that this process will converge to a steady state in which the proportion of the original 1,000 who are employed is 0.80 and the proportion who are unemployed is 0.20. We can determine that this result holds true by treating this model as a pair of difference equations.

The proportion (probability) employed in $t = 1$ is e_1. We know that the proportion employed in $t + 1$ is the proportion who had jobs in the previous period and kept them $[(1 - \alpha)e_t]$ plus the proportion who were unemployed but found jobs (βu_t). Thus the probability of employment in $t + 1$ is

$$e_{t+1} = (1 - \alpha)e_t + \beta u_t$$

Similarly the proportion unemployed in $t+1$ is the proportion who were unemployed in the previous period and stayed unemployed $[(1-\beta)u_t]$ plus the proportion who had jobs but lost them (αe_t). Thus the probability of unemployment in $t+1$ is

$$u_{t+1} = \alpha e_t + (1-\beta)u_t$$

These two difference equations make up a homogeneous system of linear difference equations that describes how the probabilities of being in the state of employment or unemployment following the initial layoff change from one period to the next. Let us solve this system.

Write the system in matrix form

$$\begin{bmatrix} e_{t+1} \\ u_{t+1} \end{bmatrix} = \begin{bmatrix} 1-\alpha & \beta \\ \alpha & 1-\beta \end{bmatrix} \begin{bmatrix} e_t \\ u_t \end{bmatrix}$$

The determinant of the coefficient matrix is $(1-\alpha)(1-\beta)-\alpha\beta$, which simplifies to $1-\beta-\alpha$. Therefore the roots (eigenvalues) of the coefficient matrix are

$$r_1, r_2 = \frac{2-\alpha-\beta}{2} \pm \frac{1}{2}\sqrt{(2-\alpha-\beta)^2 - 4(1-\beta-\alpha)}$$

Carrying out the square under the root sign and simplifying gives

$$r_1, r_2 = \frac{2-\alpha-\beta}{2} \pm \frac{1}{2}\sqrt{(\alpha+\beta)^2}$$

which simplifies further to

$$r_1, r_2 = 1 - \frac{\alpha+\beta}{2} \pm \frac{\alpha+\beta}{2}$$

The roots are therefore 1 and $1-(\alpha+\beta)$. The solutions to the homogeneous system of difference equations then are

$$e_t = C_1 + C_2(1-\alpha-\beta)^t$$

$$u_t = \frac{\alpha}{\beta}C_1 - C_2(1-\alpha-\beta)^t$$

To complete the solution, we must use the initial conditions to determine the values of the constants C_1 and C_2. At $t = 0$, e_0 and u_0 are given. Using these, we obtain

$$e_0 = C_1 + C_2$$
$$u_0 = \frac{\alpha}{\beta}C_1 - C_2$$

The first equation implies that $C_1 = e_0 - C_2$. Substitute this into the second equation to get

$$u_0 = \frac{\alpha}{\beta}(e_0 - C_2) - C_2$$

Solving this for C_2 gives

$$C_2 = \frac{\alpha e_0 - \beta u_0}{\alpha + \beta}$$

Substitute this back into the expression for C_1 to get

$$C_1 = e_0 - \frac{\alpha e_0 - \beta u_0}{\alpha + \beta}$$

Simplify this equation and use the fact that $e_0 + u_0 = 1$. This gives

$$C_1 = \frac{\beta}{\alpha + \beta}$$

The solutions to the difference equations then become

$$e_t = \frac{\beta}{\alpha + \beta} + \frac{\alpha e_0 - \beta u_0}{\alpha + \beta}(1 - \alpha - \beta)^t$$
$$u_t = \frac{\alpha}{\alpha + \beta} - \frac{\alpha e_0 - \beta u_0}{\alpha + \beta}(1 - \alpha - \beta)^t$$

The solutions give the probabilities of being in the state of employment or unemployment during any time period t. However, our real concern is to determine whether these probabilities converge to stationary values. Inspection of the solutions indicates that this is indeed the case. Because α and β are both positive but less than 1, it is necessarily true that the term $(1-\alpha-\beta)$ is between -1 and 1. Thus this term raised to the power of t goes to 0 as $t \to \infty$. As a result we conclude that

$$\lim_{t\to\infty} e_t = \frac{\beta}{\alpha+\beta}$$

$$\lim_{t\to\infty} u_t = \frac{\alpha}{\alpha+\beta}$$

In the numerical example given above, $\beta = 0.2$ and $\alpha = 0.05$. Therefore we conclude that the proportion of the original 1,000 workers who are employed during any time period converges to 80%, or 800 workers. Another 20%, or 200 workers, will be unemployed at any one time. Of course, there will continue to be transitions of people between the two states of employed and unemployed so the makeup of the two groups continues to change, but the proportion in each group does tend to converge to these values.

PRACTICE EXERCISES

S24.1. Consider the following nonlinear market equilibrium model. Price adjusts to excess demand as given by

$$\dot{p} = \alpha(q^D - q^S)$$

and the number of firms in the industry adjusts to excess profits according to

$$\dot{N} = \gamma(p - \bar{c})$$

where $Q^D = a + bp$ is the demand function and the supply function is

$$Q^s = (F + Gp)N$$

S24.2. Suppose that the daily sales of an ice-cream vendor can be classified as either high or low. Suppose that a day of high sales is followed by another day of high sales 75% of the time, and is followed by a day of low sales 25% of the time. A day of low sales is followed by another day of low sales 50% of the time and by high sales 50% of the time. Construct the system of difference equations for this Markov chain model and solve. Find the limiting values of the probability that the vendor will experience a day of high sales and a day of low sales.

S24.3. Suppose that a political party is selecting a new leader from two candidates, A and B. The winner will be decided by vote using a $2/3$ majority rule. That means that a candidate must obtain at least $2/3$ of the votes before being declared a winner. If neither candidate obtains $2/3$ of the votes, a second round of voting occurs. Party members vote again in the second round and may vote differently if they wish. If a winner is still not found, a third round of voting occurs. This process continues until one of the candidates obtains $2/3$ of the votes.

Given the probability that a voter who voted for A in the previous round will vote for A again is 0.85 and the probability that a voter who voted for B in the previous round will vote for B again is 0.90, will a $2/3$ majority ever be reached if the votes are split equally between the two candidates in the first round of voting?

Solutions

S24.1. After making appropriate substitutions, the nonlinear differential equation system is

$$\begin{aligned} \dot{p} &= \alpha a + \alpha b p - \alpha N(F + Gp) \\ \dot{N} &= \gamma(p - \bar{c}) \end{aligned}$$

The coefficient matrix of the linearized form is

$$A = \begin{bmatrix} \alpha b - \alpha G N & -\alpha(F + Gp) \\ \gamma & 0 \end{bmatrix}$$

The determinant is positive and the trace is negative. Therefore the steady state is stable. The roots are

$$\begin{aligned} r_1,\ r_2 = {} & \frac{\alpha b - \alpha G N}{2} \\ & \pm \frac{1}{2}\sqrt{(\alpha b - \alpha G N)^2 - 4\alpha\gamma(F + G\bar{c})} \end{aligned}$$

The term under the square root sign is positive as γ gets smaller (real-valued roots, indicating a stable node) but could become negative as γ gets larger (complex-valued roots, indicating a stable focus).

S24.2. Let p_t be the probability of high sales and q_t be the probability of low sales.

$$\begin{aligned} p_{t+1} &= 0.75 p_t + 0.5 q_t \\ q_{t+1} &= 0.25 p_t + 0.5 q_t \end{aligned}$$

The roots of this system are $r_1,\ r_2 = 1,\ 0.25$. The solutions then are

$$p_t = C_1 + C_2(0.25)^t$$

$$q_t = \frac{C_1}{2} - C_2(0.25)^t$$

At $t = 0$ we have

$$p_0 = C_1 + C_2$$

$$q_0 = \frac{C_1}{2} - C_2$$

Solving gives

$$C_2 = \frac{p_0 - 2q_0}{3}$$

$$C_1 = \frac{2}{3}(p_0 + q_0)$$

However, since $p_0 + q_0 = 1$, we have $C_1 = 2/3$. The solutions become

$$p_t = \frac{2}{3} + \frac{p_0 - 2q_0}{3}(0.25)^t$$

$$q_t = \frac{1}{3} - \frac{p_0 - 2q_0}{3}(0.25)^t$$

The limiting value of p_t as t goes to infinity is $2/3$; the limit of q_t is $1/3$.

S24.3. Using the notation of the Markov model of layoff in the chapter, we have $\alpha = 0.15$ and $\beta = 0.10$ here. Let p_t be the probability that a voter votes for A; let q_t be the probability that a voter votes for B. The solutions, using $p_0 = q_0 = 1/2$, are

$$p_t = 0.4 + 0.1(0.75)^t$$
$$q_t = 0.6 - 0.1(0.75)^t$$

We see that p_t begins at $p_0 = 0.5$ and then declines monotonically and converges to 0.4. Likewise q_t begins at $q_0 = 0.5$ and then rises monotonically toward its limiting value of 0.6. A 2/3 majority is never reached.

Chapter S25 Optimal Control Theory

Contents

S25.1 A Derivation of the Necessary Conditions in Optimal Control Theory

Why are the conditions that make up the maximum principle combined with the transversality condition necessarily satisfied only along the path that solves the dynamic optimization problem? In an effort to answer this question and at the same time justify these necessary conditions, we now demonstrate how it is possible to derive these necessary conditions using basic tools of calculus.

The problem at hand is to find the path of $y(t)$ that maximizes the functional J in definition 25.1. If we knew nothing about optimal control theory, how would we go about this? Our approach is to transform this maximization problem into one for which the standard rules of calculus apply. In so doing, we will demonstrate the validity of the maximum principle of optimal control theory.

Form the following expression:

$$J = \int_0^T (f[x(t), y(t), t] + \lambda(t)\{g[x(t), y(t), t] - \dot{x}\})\, dt \qquad \text{(S25.1)}$$

In forming this expression, we have added the term $\{g[x(t), y(t), t] - \dot{x}\}$ multiplied by an arbitrary function of time $\lambda(t)$. Since this term is always equal to 0 provided the constraint is satisfied at each point in time, it does not alter the value of the objective functional, J. We can think of this new expression as a Langrangean function, which makes $\lambda(t)$ a sequence or path of Lagrange multipliers as a function of t.

Next, we can actually carry out some of the integration in equation (S25.1) before attempting to maximize J. The term $\int \lambda(t)\dot{x}\,dt$ can be integrated by parts as follows. First, recall the rule for integrating by parts (section 16.5)

$$\int_0^T v\,du = [uv]_0^T - \int_0^T u\,dv$$

Let $v = \lambda(t)$ and $du = \dot{x}\,dt$. Differentiating v gives

$$dv = \dot{\lambda}(t)\,dt$$

and integrating du gives $u = x(t)$.

We therefore have

$$\begin{aligned}\int_0^T \lambda(t)\dot{x}\,dt &= [x(t)\lambda(t)]_0^T - \int_0^T x(t)\dot{\lambda}\,dt \\ &= x(T)\lambda(T) - x(0)\lambda(0) - \int_0^T x(t)\dot{\lambda}\,dt\end{aligned}$$

Substituting this into equation (S25.1) gives

$$\begin{aligned}J = &\int_0^T \{f[x(t), y(t), t] + \lambda(t)g[x(t), y(t), t] + x(t)\dot{\lambda}\}\,dt \\ &- x(T)\lambda(T) + x(0)\lambda(0)\end{aligned} \tag{S25.2}$$

To simplify notation, let us define the first two terms under the integral sign as a new function H:

$$H[x(t), y(t), \lambda(t), t] = f[x(t), y(t), t] + \lambda(t)g[x(t), y(t), t] \tag{S25.3}$$

where H depends on $x(t)$, $y(t)$, $\lambda(t)$ and t.

Using this definition of H, our objective functional to be maximized in equation (S25.2) becomes

$$J = \int_0^T \{H[x(t), y(t), \lambda(t), t] + x(t)\dot{\lambda}\}\,dt - x(T)\lambda(T) + x(0)\lambda(0) \tag{S25.4}$$

To this point, we have merely found a new way of expressing the value of the objective functional we wish to maximize. Now suppose that there is a known solution path $y^*(t)$ that maximizes J in equation (S25.4), and let $x^*(t)$ be the associated solution for the state variable. We therefore have

$$J^* = \int_0^T \{H[x^*(t), y^*(t), \lambda(t), t] + x^*(t)\dot{\lambda}\}\, dt - x^*(T)\lambda(T) + x(0)\lambda(0) \quad \text{(S25.5)}$$

which is the maximum value of J that can be attained. Since, by assumption, the path $y^*(t)$ maximizes J, we know that any other path, $y(t)$, will yield a smaller value of J. Consider any arbitrary neighboring path to $y^*(t)$. We know that if we allowed this neighboring path to get closer and closer to the optimal path $y^*(t)$, the value of J would get closer and closer to the maximum value J^*. This gives us a clue about how to find a unique property that will be true of the optimal path but not true for any neighboring paths. A convenient way to generate a family of arbitrary neighboring paths is to make use of a **perturbing path,** $z(t)$, which can be any arbitrary continuous path over time. Figure S25.1 displays both a hypothetical optimal path, $y^*(t)$, and a hypothetical perturbing path $z(t)$.

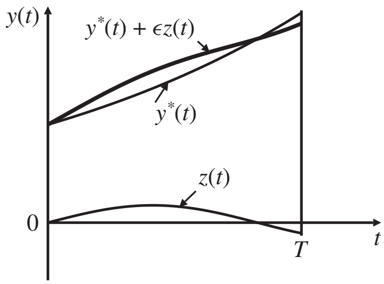

Figure S25.1 Perturbing path, $z(t)$, used to create neighboring paths to the optimal path

By adding the amount $\epsilon z(t)$ to $y^*(t)$, where ϵ is a small number, we can generate a neighboring path to $y^*(t)$ as shown in figure S25.1. The equation for the neighboring path is

$$y(t) = y^*(t) + \epsilon z(t)$$

By treating $y^*(t)$ and $z(t)$ as given paths, we can generate an entire family of neighboring paths by simply altering the value of the small constant ϵ. Varying ϵ will at the same time generate a family of neighboring paths of the state variable $x(t)$, via the equation of motion $\dot{x} = g[x(t), y(t), t]$. Since it is not possible to derive an explicit expression for the neighboring paths of $x(t)$ as ϵ is varied, we simply write the state path associated with a particular value of ϵ as $x(t, \epsilon)$ with the properties

$$x(t, 0) = x^*(t)$$
$$x(0, \epsilon) = x_0$$

The first property states that when $\epsilon = 0$, the neighboring path is, in fact, the optimal path itself. The second property states that all neighboring state paths satisfy the initial condition for the state variable.

The value of the objective functional for neighboring paths generated by varying the size of ϵ can be written as a function of ϵ:

$$J(\epsilon) = \int_0^T \{H[x(t, \epsilon), y^*(t) + \epsilon z(t), t] + x(t, \epsilon)\dot{\lambda}\}\, dt - x(T, \epsilon)\lambda(T) + x(0)\lambda(0) \quad \text{(S25.6)}$$

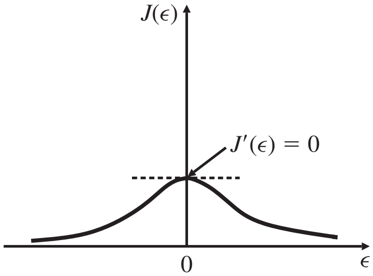

Figure S25.2 $J(\epsilon)$ maximized at $\epsilon = 0$

Since $y^*(t)$ and $z(t)$ are fixed paths, we can, by varying the size of ϵ, vary the value of the objective function J. We know that $J(\epsilon) < J^*$ for $\epsilon \neq 0$ and that $J(0) = J^*$. In other words, $J(\epsilon)$ is maximized at $\epsilon^* = 0$. As figure S25.2 demonstrates, $J'(\epsilon^*) = 0$ where $\epsilon^* = 0$. Thus, if we take the derivative of equation (S25.6) with respect to ϵ and set it equal to zero, we should obtain a first-order condition that provides a derivative property that holds only on the optimal path. This property can later be used to help identify optimal paths.

Setting $J'(\epsilon) = 0$ gives

$$J'(\epsilon) = \int_0^T \left[\frac{\partial H}{\partial x} x_\epsilon(t, \epsilon) + \frac{\partial H}{\partial y} z(t) + \dot{\lambda} x_\epsilon(t, \epsilon)\right] dt - x_\epsilon(T, \epsilon)\lambda(T) = 0 \tag{S25.7}$$

where $x_\epsilon(t, \epsilon) \equiv \partial x(t, \epsilon)/\partial \epsilon$, and where, for convenience, we have not written out the arguments of H. Collecting terms in (S25.7) gives

$$J'(\epsilon) = \int_0^T \left\{\frac{\partial H}{\partial y} z(t) + \left[\frac{\partial H}{\partial x} + \dot{\lambda}\right] x_\epsilon(t, \epsilon)\right\} dt - x_\epsilon(T, \epsilon)\lambda(T) = 0 \tag{S25.8}$$

This expression contains three terms, two of them under the integral sign. By chance the terms could cancel one another and equal zero, as required, for some specific perturbing curve $z(t)$; however, we know that $J'(\epsilon)$ must equal zero for any arbitrary perturbing curve. Therefore, each of the three terms must be identically zero. Because the perturbing curves $z(t)$ and $x_\epsilon(t, \epsilon)$ for $t \in (0, T)$ are not zero, the only way for this to happen is for each of the following three conditions to hold:

$$\frac{\partial H}{\partial y} = 0 \tag{S25.9}$$

$$\frac{\partial H}{\partial x} + \dot{\lambda} = 0 \tag{S25.10}$$

$$\lambda(T) = 0 \tag{S25.11}$$

These three conditions are necessary to make $J'(\epsilon) = 0$. Since this occurs only when $\epsilon = 0$, these conditions therefore hold *only* on the optimal paths $y^*(t)$ and $x^*(t)$. As a result, they are necessary conditions for the maximization of the objective functional J.

For completeness, we add that another condition that must hold along the optimal path is

$$\frac{\partial H}{\partial \lambda} = \dot{x} - g[x^*(t), y^*(t), t] = 0 \tag{S25.12}$$

This condition merely recovers the constraint that was specified in the maximization problem. We include it here as a necessary condition because we had to assume this constraint holds for all $t \in (0, T)$ in order to go from equation (25.1) to equation (S25.1).

S25.2 Interpretation of λ

The interpretation of λ as the shadow price or imputed marginal value of the state variable is obtained by differentiating equation (S25.5) (an expression for J^*, the maximum value function) with respect to $x(0)$. This gives

$$\frac{\partial J^*}{\partial x(0)} = \lambda(0)$$

This says that a marginal increase in $x(0)$ (the initial value of the state variable, e.g., the capital stock) increases the maximum value of the objective (such as profits) by an amount equal to $\lambda(0)$. Hence $\lambda(0)$ is the marginal imputed value or shadow price of $x(0)$. Similarly differentiating equation (S25.5) with respect to $x(T)$ gives

$$\frac{\partial J^*}{\partial x(T)} = -\lambda(T)$$

from which we conclude that a marginal increase in the amount of the state variable that must be left over at the end of the planning horizon decreases the maximum value of the objective by $\lambda(T)$. Hence $\lambda(T)$ is the marginal imputed value or shadow price of $x(T)$. It follows intuitively that $\lambda(t)$ can be interpreted as the marginal imputed value or shadow price of $x(t)$, but we do not attempt to prove this result here.

S25.3 Derivation of the $H(T) = 0$ Condition

To understand the origin of definition 25.7, consider the expression we derived for J in equation (S25.5), but which we now write as a function of T

$$J(T) = \int_0^T \{H[x^*(t), y^*(t), \lambda(t), t] + x^*(t)\dot{\lambda}\}\, dt - x^*(T)\lambda(T) + x(0)\lambda(0)$$

Take the derivative of this expression with respect to T. This gives

$$\begin{aligned} J'(T) = & \int_0^T \left\{ \left[\frac{\partial H}{\partial x} + \dot{\lambda} \right] \frac{dx^*(t)}{dT} + \frac{\partial H}{\partial y} \frac{dy^*(t)}{dT} \right\} dt \\ & + H[x^*(T), y^*(T), T] + x^*(T)\dot{\lambda}(T) - \frac{d[x^*(T)\lambda(T)]}{dT} \\ = & \ 0 \end{aligned} \tag{S25.13}$$

To understand this, you need to use Leibniz's rule, which was developed in section 16.5. Suppose that y is given by

$$y = \int_L^U f(t, z)\, dt$$

That is, y is equal to the integral over t, where t runs from L up to U, of a function that depends on t and a variable z that itself depends on x. The derivative of y with respect to x is

$$\frac{dy}{dx} = \int_L^U \frac{\partial f(t, z)}{\partial z} \frac{dz}{dx}\, dt + f(U, z) \frac{dU}{dx} - f(L, z) \frac{dL}{dx}$$

The derivative is defined as: the integral of the derivative of f with respect to x plus the integrand, evaluated at U, multiplied by the derivative of U, minus the integrand, evaluated at L, multiplied by the derivative of L. If U and L are not functions of x, then the last two terms are zero. If the variable z is not a function of x, then the first term is zero. In the derivative of J with respect to T, the derivative of the upper limit of integration with respect to T is just one, but the lower limit is not a function of T.

In equation (S25.13), the necessary conditions we have already derived imply that the integral term is zero along an optimal path. After carrying out the differentiation of the term $[x^*(T)\lambda(T)]$ then, equation (S25.13) becomes

$$J'(T) = H[x^*(T), y^*(T), T] + x^*(T)\dot{\lambda}(T) - x^*(T)\dot{\lambda}(T) - \lambda(T)\dot{x}(T) = 0$$

where the derivatives of $\lambda(T)$ and $x(T)$ with respect to T are denoted by a dot over the variable. Since $\lambda(T) = 0$ is another necessary condition, this reduces to

$$J'(T) = H[x^*(T), y^*(T), T] = 0$$

This is the rationale for the additional necessary condition when T is free to be chosen optimally. The derivation says that the change in the optimal value function if T is changed marginally is equal to the value of the maximized Hamiltonian at time T.

PRACTICE EXERCISE

S25.1. In the following optimal investment problem, suppose that there is a *salvage value* for the firm's capital stock. That means that at the end of the planning horizon, the firm's capital stock has a salvage value given by the function $v[K(T)]$, with $v' > 0$ and $v'' < 0$. This could represent, for example, the amount for which the firm could sell its capital in the second-hand market. The problem is as follows:

$$\max \int_0^T e^{-\rho t}\Pi(K, I)\, dt + e^{-\rho T} v[K(T)]$$

$$\begin{aligned} \text{subject to} \quad & \dot{K} = I - \delta K \\ & K(0) = K_0 > 0 \\ & K(T) \geq 0 \\ & T \qquad \text{(free)} \end{aligned}$$

Derive the necessary conditions for this problem using the method that begins with equation (S25.1). Prove that the only changes are to the transversality conditions and that these become

$$\mu(T) = v'[K(T)] \quad \text{if } K(T) \geq 0; \quad K(T) = 0 \quad \text{otherwise}$$
$$\mathcal{H}(T) = \rho v[K(T)] \quad \text{if } T \text{ finite}$$

Provide an economic interpretation of these transversality conditions.

Solution

S25.1.

$$J = \int_0^T \{e^{-\rho t}\Pi(K, I) + \lambda(t)[I - \delta K - \dot{K}]\}dt + e^{-\rho T} V(K(T))$$

Integrating $\lambda \dot{K}$ by parts as shown in the chapter gives

$$J = \int_0^T \{e^{-\rho t}\Pi(K, I) + \lambda(t)[I - \delta K] + K(t)\dot{\mu}\}dt$$
$$-\lambda(T)K(T) + \lambda(0)K_0 + e^{-\rho T}V(K(T))$$

Define

$$\mathcal{H} = \{e^{-\rho t}\Pi(K, I) + \lambda(t)[I - \delta K]\}e^{\rho t}$$

and define $\mu(t) = \lambda(t)e^{\rho t}$ so

$$\mathcal{H} = \Pi(K, I) + \mu(t)[I - \delta K]$$

Making these substitutions gives

$$J = \int_0^T \{e^{-\rho t}\mathcal{H}(K(t), I(t)) + K(t)(\dot{\mu} - \rho\mu)e^{-\rho t}\}dt$$
$$-\mu(T)K(T)e^{-\rho T} + \mu(0)K_0 + e^{-\rho T}V(K(T))$$

Now use the perturbing path to get

$$J(\epsilon) = \int_0^{T(\epsilon)} e^{-\rho t}\{\mathcal{H}[K(t, \epsilon), I^* + \epsilon z(t)] + K(t, \epsilon)(\dot{\mu} - \rho\mu)\}dt$$
$$-\mu(T)K(T, \epsilon)e^{-\rho T} + \mu(0)K_0 + e^{-\rho T}V(K(T, \epsilon))$$

$$J'(\epsilon) = \int_0^{T(\epsilon)} e^{-\rho t}\left\{\left[\frac{\partial \mathcal{H}}{\partial K} + \dot{\mu} - \rho\mu\right]K_\epsilon(t, \epsilon) + \frac{\partial \mathcal{H}}{\partial I}z(t)\right\}dt$$
$$-\mu(T)K_\epsilon(T, \epsilon)e^{-\rho T} + e^{-\rho T}V'(K(T))K_\epsilon(T, \epsilon) + \frac{\partial J(\epsilon)}{\partial T} = 0$$

Leaving $\partial J/\partial T$ aside for the moment, the necessary conditions become

$$\frac{\partial \mathcal{H}}{\partial K} + \dot{\mu} - \rho\mu = 0$$

$$\frac{\partial \mathcal{H}}{\partial I} = 0$$

$$-\mu(T) + V'(K(T)) = 0 \quad \text{if } T \text{ is finite}$$

Next, using equation (25.65) in the chapter, we see that $\partial J/\partial T$ reduces to the following after using the above three conditions:

$$\frac{\partial J}{\partial T} = \mathcal{H}[K^*(T), I^*(T)]e^{-\rho T} + K^*(T)[\dot{\mu}(T) - \rho\mu(T)]e^{-\rho T}$$
$$-\dot{\mu}(T)K^*(T)e^{-\rho T} + \rho\mu(T)K^*e^{\rho T} - \rho e^{-\rho T}V(K^*(T)) = 0$$

which reduces to

$$\mathcal{H}[K^*(T), I^*(T)] - \rho V(K^*(T)) = 0 \quad \text{if } T \text{ is finite}$$

The first of the new conditions requires the shadow price of $K(T)$ to equal the marginal salvage value of $K(T)$. The second new condition says to stop operating the firm at T when the flow of economic profits from operating has just become equal to the flow of economic profits from not operating (living off the interest from the salvage value).

Appendix Complex Numbers and Circular Functions

In chapter 20 we learned that it is necessary to find the roots of the characteristic equation

$$r^2 + a_1 r + a_2 = 0$$

to solve second-order linear difference equations. In chapter 23 the characteristic roots are required in the solution of second-order linear differential equations, and in chapter 24 they are required to solve *systems* of linear difference and differential equations. The characteristic roots are given by

$$r_1, r_2 = \frac{-a_1}{2} \pm \frac{1}{2}\sqrt{a_1^2 - 4a_2}$$

When $a_1^2 - 4a_2 > 0$, r_1 and r_2 are real-valued numbers. However, when $a_1^2 - 4a_2 < 0$, there are no real valued numbers that solve the characteristic equation. Although this may appear to make it impossible to go on to solve the difference equation, it turns out that there is a way to solve this problem: make use of the complex valued number system. This system, combined with a powerful theorem (de Moivre's theorem in the case of difference equations and Euler's theorem in the case of differential equations) allows us to find real valued solutions to the difference (or differential) equation even though there are no real-valued roots to the characteristic equation. The purpose of this appendix is to explain how to do this. We begin with an introduction to the concept of complex numbers.

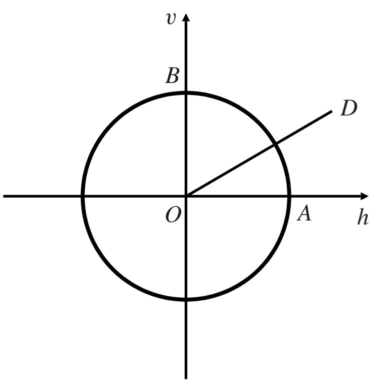

Figure A.1 Argand diagram

Complex Numbers

A complex number can be defined as an ordered pair of real numbers (h, v). For example, $(2, 1)$ and $(1, 2)$ are both complex numbers, though different ones because the order matters. A complex number is a two-dimensional concept but can be thought of as a single entity with the help of a graphical representation. Figure A.1, called an **Argand diagram,** shows the complex number as the vector OD where D is represented by its Cartesian coordinates, (h, v).

There are a number of rules in algebra for this number system. Two that are of immediate interest to us are *addition* and *scalar multiplication*.

Addition of two complex numbers $z_1 = (h_1, v_1)$ and $z_2 = (h_2, v_2)$ is carried out as follows:

$$z_1 + z_2 = (h_1 + h_2, v_1 + v_2)$$

giving us a new complex number.

Multiplication of a complex number by a real-valued scalar a yields

$$az_1 = (ah_1, av_1)$$

Two complex numbers in particular receive special attention. These are

$$(1, 0) = \text{unit vector } OA \quad \text{and} \quad (0, 1) = \text{unit vector } OB$$

The unit vector $(1, 0)$ can be thought of as having the real value 1 (the reason for this is made clear below). The other unit vector, $(0, 1)$, is referred to as i. As we will see, this i is the imaginary number that has the property $i = \sqrt{-1}$.

Given these definitions, scalar multiplication of the unit vector one and the unit vector i yields:

$$a1 = a(1, 0) = (a, 0) \quad \text{and} \quad ai = a(0, 1) = (0, a)$$

As a result these two vectors provide us with a convenient way of representing any complex number (h, v) as $h + vi$:

$$h + vi = h(1, 0) + v(0, 1) = (h, 0) + (0, v) = (h + 0, 0 + v) = (h, v)$$

Thus a complex number z can be expressed as

$$z = h + vi$$

where h and v are real numbers and i is the imaginary number defined above. The set of complex numbers is larger than and includes the set of real numbers. The set of real numbers could be defined as the subset of complex numbers that occurs when $v = 0$.

The rationale for giving the unit vector $(1, 0)$ the real value of unity and the unit vector i the value $\sqrt{-1}$ follows logically from the rules of vector multiplication. It would take us too far afield to explain the rules for multiplying two complex numbers together; however, we will state two useful results.

First, any vector when multiplied by the unit vector $(1, 0)$ remains the same vector. For this reason we say that the unit vector $(1, 0)$ has the value one. Second, any vector, when multiplied by the unit vector $i = (0, 1)$, is rotated 90° in the counterclockwise direction. Thus the unit vector $(1, 0)$, when multiplied by the

i vector, becomes the i vector. Similarly the i vector, when multiplied by itself, becomes the $(0, -1)$ vector, which has the real value -1. It is for this reason that we use the rule that $i^2 = -1$, which leads to $i = \sqrt{-1}$.

Two additional properties of complex numbers that are used in chapters 20, 23, and 24 are

Theorem A.1 The sum of two conjugate complex numbers is a real-valued number.

Proof

Two conjugate complex numbers are $z_1 = (h, v)$ and $z_2 = (h, -v)$. Adding them gives us

$$(h + vi) + (h - vi) = 2h$$

a real-valued number. ■

Theorem A.2 The difference between two conjugate complex numbers multiplied by i yields a real-valued number.

Proof

$$[(h + vi) - (h - vi)]i = 2vi^2 = -2v$$

a real-valued number. ■

Now that we have introduced the concept of complex numbers, we need to show how the concept helps solve the difference (or differential) equation when $a_1^2 - 4a_2 < 0$. In particular, we need to know how to interpret the expressions

$$(h + vi)^t \quad \text{and} \quad (h - vi)^t$$

in the case of difference equations and

$$e^{(h+vi)t}$$

and

$$e^{(h-vi)t}$$

in the case of differential equations. It turns out that to do this, we need to re-express the complex number (h, v) in terms of its *polar coordinates*. Therefore we turn next to a brief review of *circular* (or *trigonometric*) functions.

Circular Functions

The circular functions are best explained with the aid of a diagram containing a circle with a unit radius. Figure A.2 shows the unit circle centered on the origin. The point marked A is $(1, 0)$, the point marked B is $(0, 1)$, A' is $(-1, 0)$, and B' is $(0, -1)$. The vector OP is also a radius of unit length and it makes an angle of θ degrees. As P moves around the circle in the counterclockwise (positive) direction, the angle takes on values ranging from $0°$ to $360°$. For example, at B the angle is $90°$, at A' it is $180°$, at B' it is $270°$, and back at A after one revolution it is $360°$.

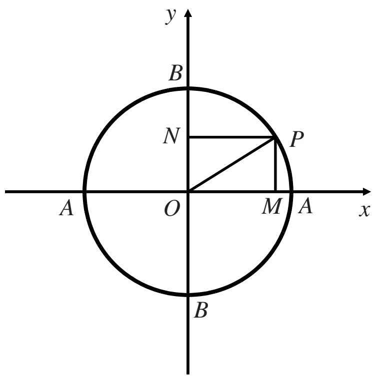

Figure A.2 Unit circle

Drop a perpendicular from P to the horizontal axis at M. Notice that as P moves around the circle, the length of OM changes in value. For example, at B, the length of OM is zero; at A', OM is -1; at B', OM is zero again, and back at A after one revolution, OM is one. Thus there is an explicit relationship between the angle θ made by OP and the distance OM. We define this as the cosine relationship and give it the formal definition

$$\cos\theta° = \frac{OM}{OP} = OM$$

For example, we have just seen that $\cos(0°) = 1$, $\cos(90°) = 0$, $\cos(180°) = -1$, $\cos(270°) = 0$, and $\cos(360°) = 1$.

A similar process generates the sine function. Extend a perpendicular from P to the vertical axis at N. Again, as P moves around the circle in the counterclockwise direction, $0N$ takes on values starting with zero at A, rising to one at B, falling to zero at A', falling further to negative one at B' and rising back to zero at A. Formally we have

$$\sin\theta° = \frac{ON}{OP} = ON$$

We can see from the unit circle that some particular values of the sine function are $\sin(0°) = 0$, $\sin(90°) = 1$, $\sin(180°) = 0$, $\sin(270°) = -1$, and $\sin(360°) = 0$.

Although we are all accustomed to measuring angles in degrees, it is actually easier (and customary) in theoretical work with angles to measure them in terms of the distance of the arc AP taken counterclockwise. The units of distance are called *radians*. We can calculate the relationship between 1 radian and $1°$ easily because we know the distance around the entire circumference of the circle is $2\pi R$ where $R = 1$ is the radius of the unit circle, giving us a distance of 2π. Since there are 360° in one complete revolution of the circle and a distance of 2π radians, we know that

$$\begin{aligned} 360^\circ &\text{ corresponds to } 2\pi \text{ radians} \\ 1^\circ &\text{ corresponds to } \frac{\pi}{180} \text{ radians} \\ x^\circ &\text{ corresponds to } x\frac{\pi}{180} \text{ radians} \end{aligned}$$

Angles larger than $360^\circ = 2\pi$ radians are obtained by letting P rotate more than once around the circle. For example, starting at P and going once around gives us an angle of $\theta^\circ + 360^\circ$, twice around gives us an angle of $\theta^\circ + 720^\circ$, and so on. In terms of radians, if the angle is x radians, once around gives us an angle of $x + 2\pi$ radians, twice around gives us an angle of $x + 4\pi$ radians, and so on. In all of the following, we use radian measures of angles.

The circular functions are defined as before but now using radians. If $\theta^\circ = x$ radians, then $\sin(x) = ON$ and $\cos(x) = OM$. Some commonly used values of the sine and cosine functions are

$$\begin{aligned} 0^\circ &= 0 \text{ radians} : \sin 0 = 0; \cos 0 = 1 \\ 90^\circ &= \frac{\pi}{2} \text{ radians} : \sin\left(\frac{\pi}{2}\right) = 1; \cos\left(\frac{\pi}{2}\right) = 0 \\ 180^\circ &= \pi \text{ radians} : \sin(\pi) = 0; \cos(\pi) = -1 \\ 270^\circ &= \frac{3\pi}{2} \text{ radians} : \sin\left(\frac{3\pi}{2}\right) = -1; \cos\left(\frac{3\pi}{2}\right) = 0 \\ 360^\circ &= 2\pi \text{ radians} : \sin(2\pi) = 0; \cos(2\pi) = 1 \end{aligned}$$

Figure A.3 shows the sine function. We see that $\sin(x) = ON$ rises from zero to one as x increases from zero to $\pi/2$, then decreases back to zero as the angle x increases from $\pi/2$ to π. It then decreases to negative one as x increases further to $3\pi/2$ and then rises back to zero as x increases to 2π. There is a whole cycle of $\sin(x)$ going from zero to one, back to zero and then to negative one before returning to zero. The cycle repeats itself as P makes a second revolution around the circle and angle x continues to rise.

The cosine function is shown in figure A.4. Referring back to figure A.2 we see that $\cos x = OM$ begins at one when $x = 0$, then falls to zero as x increases to $\pi/2$, then falls further to negative one, then rises back up to zero, and then completes the first cycle by returning to its starting position at one as x increases through π, $3\pi/2$, and 2π. As does the sine function, the cosine function repeats its cycle every 2π radians.

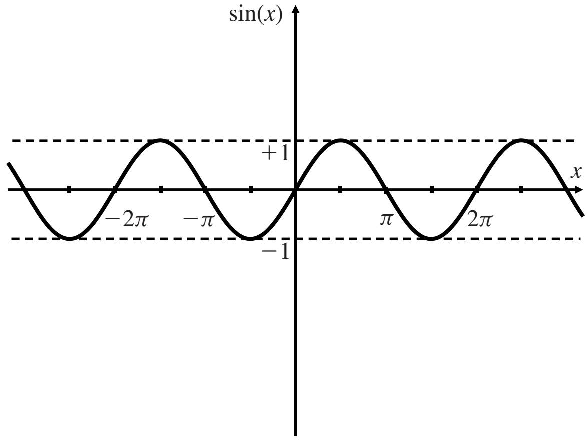

Figure A.3 Sine function

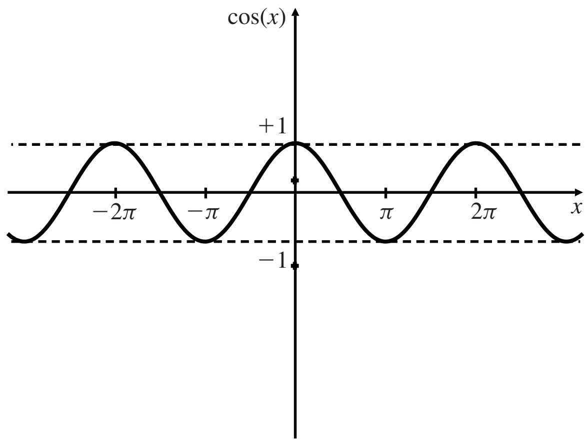

Figure A.4 Cosine function

We can see now how the circular functions generate cyclical behavior. Both the sine and cosine functions repeat themselves every 2π radians; they are said to have a period of 2π radians. This property means that $\sin(x) = \sin(x + 2\pi) = \sin(x + 4\pi) = \ldots$ and $\cos(x) = \cos(x + 2\pi) = \cos(x + 4\pi) = \ldots$. In general,

$$\sin(x) = \sin(x + 2n\pi) \text{ and } \cos(x) = \cos(x + 2n\pi), \qquad n = 0, 1, 2, \ldots$$

Both the sine and cosine functions are bounded between one and negative one. They are said to have an amplitude of one. They also obey some important properties we will need to use to solve difference equations and differential equations. First, because negative angles are constructed by rotating P around the circle in the clockwise, rather than the counterclockwise, direction, the sine and cosine functions are symmetrical between positive and negative values of the domain. Therefore

$$\sin(-x) = -\sin(x) \quad \text{and} \quad \cos(-x) = \cos(x)$$

Another useful property is that the functions

$$\phi_1(x) = A\sin(x) \quad \text{and} \quad \phi_2(x) = A\cos(x)$$

where A is a constant, have the same period as the sine and cosine functions (2π) but are bounded between A and $-A$ (have an amplitude of A).

The final properties of the sine and cosine functions we require are

$$\frac{d}{dx}\sin(x) = \cos(x)$$
$$\frac{d}{dx}\cos(x) = -\sin(x)$$

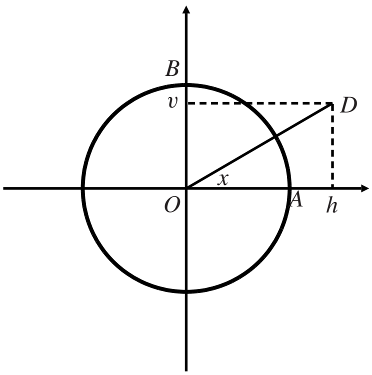

Figure A.5 Complex number in polar coordinates

An alternative representation of a complex number is one that recognizes, as in figure A.5, that the point D can be described not only by its *Cartesian coordinates* (h, v), but by the length of the ray OD and the angle, x radians, it makes with the horizontal axis.

In particular, since $\cos(x) = Oh/OD = h/R$ where R is the length of the vector OD, we know that $h = R\cos(x)$. Similarly, since $\sin(x) = Ov/OD = v/R$, we know that $v = R\sin(x)$. These results allow us to express a complex number as

$$h + vi = R\cos(x) + Ri\sin(x) = R[\cos(x) + i\sin(x)] \tag{A.1}$$

This expresses the complex number in terms of its *polar coordinates* (R, x) as opposed to its *Cartesian coordinates* (h, v). Although they are just two different ways of determining the location of point D, it turns out to be much easier in what follows to use the polar coordinates definition of a complex number.

Euler's Formula

We are now able to demonstrate that

$$e^{ix} = \cos(x) + i\sin(x)$$
$$e^{-ix} = \cos(x) - i\sin(x)$$

which are known as Euler's formulas. To do this we take Taylor series expansions of the exponential function and the sine and cosine functions around the point $x = 0$; in this form, it is straightforward to derive Euler's formula.

A Taylor series expansion of the function e^z around the point $z = 0$ gives us

$$e^z = e^0 + \frac{e^0(z-0)}{1!} + \frac{e^0(z-0)^2}{2!} + \frac{e^0(z-0)^3}{3!} + \frac{e^0(z-0)^4}{4!} + \cdots$$

Therefore

$$e^z = 1 + \frac{z}{1!} + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \frac{z^5}{5!} + \frac{z^6}{6!} + \frac{z^7}{7!} + \cdots$$

Setting $z = ix$ and recalling that $i^2 = -1$, we get

$$e^{ix} = 1 + ix - \frac{x^2}{2!} - \frac{ix^3}{3!} + \frac{x^4}{4!} + \frac{ix^5}{5!} - \frac{x^6}{6!} - \frac{ix^7}{7!} + \cdots$$

and

$$e^{-ix} = 1 - ix - \frac{x^2}{2!} + \frac{ix^3}{3!} + \frac{x^4}{4!} - \frac{ix^5}{5!} - \frac{x^6}{6!} - \frac{ix^7}{7!} + \cdots$$

The next step is to take a Taylor series expansion of the sine and cosine functions. We begin with the cosine function. Taking successive derivatives of $\cos(x)$ and evaluating them at $x = 0$ gives

$$\frac{d}{dx}\cos(x) = -\sin(x) = 0 \text{ at } x = 0$$

$$\frac{d^2}{dx^2}\cos(x) = -\cos(x) = -1 \text{ at } x = 0$$

$$\frac{d^3}{dx^3}\cos(x) = \sin(x) = 0 \text{ at } x = 0$$

$$\frac{d^4}{dx^4}\cos(x) = \cos(x) = 1 \text{ at } x = 0$$

$$\frac{d^5}{dx^5}\cos(x) = -\sin(x) = 0 \text{ at } x = 0$$

$$\frac{d^6}{dx^6}\cos(x) = -\cos(x) = -1 \text{ at } x = 0$$

$$\cdots$$

Using these derivatives in a Taylor's series expansion around $x = 0$ gives

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$$

Similarly we can derive a Taylor series expansion for $\sin(x)$

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

Using these expansions of $\cos(x)$ and $\sin(x)$, the expanded expression for e^{ix} can be rewritten to get

$$e^{ix} = i\sin(x) + \cos(x) \tag{A.2}$$

Similarly the expanded expression for e^{-ix} can be rewritten to get

$$e^{-ix} = -i\sin(x) + \cos(x) \tag{A.3}$$

This completes the derivation of Euler's formula, which is used in completing the solution to linear, second-order differential equations in the case of complex valued roots.

De Moivre's theorem shows how to interpret a complex number that is raised to the nth power.

Theorem A.3 **(De Moivre's theorem)** The conjugate complex numbers, $h \pm vi$, when raised to the nth power can be expressed as

$$(h \pm vi)^n = R^n[\cos(nx) \pm i\sin(nx)]$$

where $R = \sqrt{h^2 + v^2}$, $h = R\cos(x)$, $v = R\sin(x)$.

Proof

From equation (A.1) a conjugate complex number $h \pm vi$ can be expressed in polar coordinate form as

$$h \pm vi = R[\cos(x) \pm i\sin(x)]$$

From Euler's formula in (A.2) and (A.3), we know that

$$e^{\pm ix} = \cos(x) \pm i \sin(x)$$

Therefore

$$h \pm vi = Re^{\pm ix}$$

This means that

$$(h \pm vi)^n = R^n e^{\pm inx}$$

Applying Euler's formula again gives us the result

$$(h \pm vi)^n = R^n[\cos(nx) \pm i \sin(nx)] \qquad \blacksquare$$