INTEGRAL UNIVERSITY, LUCKNOW

DIRECTORATE OF DISTANCE EDUCATION

BBA-202 Paper Code: ME/B

MANAGERIAL ECONOMICS

CONTENTS

UNIT-1 INTRODUCTION TO ECONOMIC DECISION MAKING

CONTENTS

INTRODUCTION

In this Unit, we begin our study of managerial economics by stressing decision-making applications. In the first section, we introduce seven decision examples, all of which we will analyze in detail later in the text. Although these examples cover only some applications of economic analysis, they represent the breadth of managerial economics and are intended to whet the reader's appetite. Next, we present a basic model of the decision-making process as a framework in which to apply economic analysis. This model proposes six steps to help structure complicated decisions so that they may be clearly analyzed. After presenting the six steps, we outline a basic theory of the firm and of government decisions and objectives. In the concluding section, we present a brief overview of the topics covered in the Units to come.

SEVEN EXAMPLES OF MANAGERIAL DECISIONS

The best way to become acquainted with managerial economics is to come face to face with real-world decision-making problems. The seven examples that follow represent the different kinds of decisions that private- and public-sector managers face. All of them are revisited and examined in detail in later Units. The examples follow a logical progression. In the first example, a global carmaker faces the most basic problem in managerial economics: determining prices and outputs to maximize profit. As we shall see in Units 2 through 6, making decisions requires a careful analysis of revenues and costs. The second example highlights competition between firms, the subject of Units 7 through 10. Here, two large bookstore chains are battling for market share in a multitude of regional markets. Each is trying to secure a monopoly, but when both build superstores in the same city, they frequently become trapped in price wars.

Multinational Production and Pricing

Almost all firms face the problem of pricing their products. Consider a U.S. multinational carmaker that produces and sells its output in two geographic regions. It can produce cars in its home plant or in its foreign subsidiary. It sells cars in the domestic market and in the foreign market. For the next year, it must determine the prices to set at home and abroad, estimate sales for each market, and establish production quantities in

each facility to supply those sales. It recognizes that the markets for vehicles at home and abroad differ with respect to demand (that is, how many cars can be sold at different prices). Also, the production facilities have different costs and capacities. Finally, at a cost, it can ship vehicles from the home facility to help supply the foreign market, or vice versa. Based on the available information, how can the company determine a profit maximizing pricing and production plan for the coming year?

Market Entry

Group—engaged in a cutthroat retail battle. In major city after major city, the rivals opened superstores, often within sight of each other. By the mid-1990s, more books were sold via chain stores than by independent stores, and both companies continued to open new stores at dizzying rates. The ongoing competition raises a number of questions: How did either chain assess the profitability of new markets? Where and when should each enter new markets? What if a region's book-buying demand is sufficient to support only one superstore? What measures might be taken by an incumbent to erect entry barriers to a would-be entrant? On what dimensions—number of titles, pricing, personal service—did the companies most vigorously compete?

Building a New Bridge

As chief city planner of a rapidly growing Sun Belt city, you face the single biggest decision of your tenure: whether to recommend the construction of a new harbor bridge to connect downtown with the surrounding suburbs located on a northern peninsula. Currently, suburban residents commute to the city via a ferry or by driving a long-distance circular route. Preliminary studies have shown that there is considerable need and demand for the bridge. Indeed, the bridge is expected to spur economic activity in the region as a whole. The projected cost of the bridge is $75 million to $100 million. Part of the money would be financed with an issue of municipal bonds, and the remainder would be contributed by the state. Toll charges on commuting automobiles and particularly on trucks would be instituted to recoup a portion of the bridge's costs. But, if bridge use falls short of projections, the city will be saddled with a very expensive white elephant. What would you recommend?

SIX STEPS TO DECISION MAKING

The examples just given represent the breadth of the decisions in managerial economics. Different as they may seem, each decision can be framed and analysed using a common approach based on six steps, With the examples as a backdrop, we will briefly outline each step. Later in the text, we will refer to these steps when analyzing managerial decisions.

Step 1: Define the Problem

What is the problem the manager faces? Who is the decision maker? What is the decision setting or context, and how does it influence managerial objectives or options?

Step 2: Determine the Objective

What is the decision maker's goal? How should the decision maker value outcomes with respect to this goal? What if he or she is pursuing multiple, conflicting objectives?

Step 3: Explore the Alternatives

What are the alternative courses of action? What are the variables under the decision maker's control? What constraints limit the choice of options?

Step 4: Predict the Consequences

What are the consequences of each alternative action? Should conditions change, how would this affect outcomes? If outcomes are uncertain, what is the likelihood of each? Can better information be acquired to predict outcomes?

Step 5: Make a Choice

After all the analysis is done, what is the preferred course of action? For obvious reasons, this step (along with step 4) occupies the lion's share of the analysis and discussion in this book. Once the decision maker has put the problem in context, formalized key objectives, and identified available alternatives, how does he or she go about finding a preferred course of action?

Step 6: Perform Sensitivity Analysis

What features of the problem determine the optimal choice of action? How does the optimal decision change if conditions in the problem are altered? Is the choice sensitive to key economic variables about which the decision maker is uncertain?

PRIVATE AND PUBLIC DECISIONS: AN ECONOMIC VIEW

Our approach to managerial economics is based on a model of the firm: how firms behave and what objectives they pursue. The main tenet of this model, or **theory of the firm**, is that management strives to maximize the firm's profits. This objective is unambiguous for decisions involving predictable revenues and costs occurring during the same period of time. However, a more precise profit criterion is needed when a firm's revenues and costs are uncertain and accrue at different times in the future. The most general theory of the firm states that Management's primary goal is to maximize the value of the firm. Here, the firm's value is defined as the present value of its expected future profits. Thus, in making any decision, the manager must attempt to predict its impact on future profit flows and determine whether, indeed, it will add to the value of the firm.

Business Behavior: Maximizing Value

Value maximization is a compelling prescription concerning how managerial decisions should be made. Although this tenet is a useful norm in describing actual managerial behavior, it is not a perfect yardstick. After all, large-scale firms consist of many levels of authority and myriad decision makers. Even if value maximization is the ultimate corporate goal, actual decision making within this complex organization may look quite different. There are several reasons for this:

1. Managers may have individual incentives (such as job security, career advancement, increasing a division's budget, resources, power) that are at odds with value maximization of the total firm. For instance, it sometimes is claimed that company executives are apt to focus on short-term value maximization (increasing next year's earnings) at the expense of long-run firm value.
2. Managers may lack the information (or fail to carry out the analysis) necessary for value-maximizing decisions.
3. Managers may formulate but fail to implement optimal decisions.

Although value maximization is the standard assumption in managerial economics, three other decision models should be noted. The model of **satisficing** behavior posits that the typical firm strives for a satisfactory level of performance rather than attempting to maximize its objective. Thus, a firm might aspire to a level of annual profit, say $40 million, and be satisfied with policies that achieve this benchmark. More generally, the firm may seek to achieve acceptable levels of performance with respect to multiple objectives (profitability being only one such objective).

Public Decisions

In government decisions, the question of objectives is much broader than simply an assessment of profit. Most observers would agree that the purpose of public decisions is to promote the welfare of society, where the term society is meant to include all the people whose interests are affected when a particular decision is made. The difficulty in applying the social welfare criterion in such a general form is that public decisions inevitably carry different benefits and costs to the many groups they affect. Some groups will gain and others will lose from any public decision. In our earlier example of the bridge, businesses and commuters in the region can expect to gain, but nearby neighbors who suffer extra traffic, noise, and exhaust emissions will lose. The program to convert utilities from oil to coal will benefit the nation by reducing our dependence on foreign oil. However, it will increase many utilities' costs of producing electricity, which will mean higher electric bills for many residents. The accompanying air pollution will bring adverse health and aesthetic effects in urban areas. Strip mining has its own economic and environmental costs, as does nuclear power. In short, any significant government program will bring a variety of new benefits and costs to different affected groups.

The important question is: How do we weight these benefits and costs to make a decision that is best for society as a whole? One answer is provided by 16 Unit 1 Introduction to benefit-cost analysis, the principal analytical framework used in guiding public decisions. **Benefit-cost analysis** begins with the systematic enumeration of all of the potential benefits and costs of a particular public decision. It goes on to measure or estimate the dollar magnitudes of these benefits and costs. Finally, it follows the decision rule: Undertake the project or program if and only if its total benefits exceed its total costs. Benefit-cost analysis is similar to the profit calculation of the private firm with one key difference: Whereas the firm considers only the revenue it accrues and the cost it incurs, public decisions account for all benefits, whether or not recipients

pay for them (that is, regardless of whether revenue is generated) and all costs (direct and indirect).

Behavioral Economics

Much of economic analysis is built on a description of ultrarational self interested individuals and profit-maximizing businesses. While this framework does an admirable job of describing buyers and sellers in markets, workers interacting in organizations, and individuals grappling with major life-time decisions, we all know that real-world human behavior is much more complicated than this. The ultrarational analyzer and calculator (Mr. Spock of Star Trek) is an extreme type, a caricature.

REVIEW QUESTION

1. What is managerial economics? What role does it play in shaping business decisions?
2. Management sometimes is described as the art and science of making decisions with too little information.
3. Suppose a soft-drink firm is grappling with the decision about whether or not to introduce to the market a new carbonated beverage with 25 percent real fruit juice. How might it use the six decision steps to guide its course of action?
4. Discuss about seven examples of managerial decisions?
5. Describe about multinational production and pricing in detail.
6. Discuss about the six steps to decision making?
7. Explain about private and public decisions.

FURTHER READINGS

1. Managerial Economics - Arun Kumar, Rachana Sharma
2. Managerial Economics - Thomas J. Webster
3. Managerial Economics - Petersen / Jain
4. Managerial Economics - Yogesh Maheshwari
5. Managerial Economics - E. Narayanan Nadar, S. Vijayan

UNIT-2 OPTIMAL DECISIONS USING MARGINAL ANALYSIS

CONTENTS

INTRODUCTION

This Unit introduces the analysis of managerial decision making that will occupy us for the remainder of the book. The Unit is devoted to two man topics. The first is a simple economic model (i.e., a description) of the private, profit-maximizing firm. The second is an introduction to marginal analysis, an important tool for arriving at optimal decisions. Indeed, it is fair to say that the subsequent Units provide extensions or variations on these two themes. The present Unit employs marginal analysis as a guide to output and pricing decisions in the case of a single product line under the simplest demand and cost conditions. In Units 3 and 4, we extend marginal analysis to the cases of complex demand conditions, multiple markets, and price discrimination. In Units 5 and 6, we apply the same approach to settings that involve more complicated production technologies and cost conditions, multiple production facilities, and multiple products. In Units 7, 8, and 9, we analyze the key market environments—competition, oligopoly, and monopoly—in which the profit-maximizing firm operates. Together, these Units demonstrate the great power of marginal analysis as a tool for solving complex decisions. Consequently, it is important to master the logic of marginal analysis at the outset. We start with a simple example before turning to the model of the firm.

A SIMPLE MODEL OF THE FIRM

The decision setting we will investigate can be described as follows:

1. A firm produces a single good or service for a single market with the objective of maximizing profit.
2. Its task is to determine the quantity of the good to produce and sell and to set a sales price.
3. The firm can predict the revenue and cost consequences of its price and output decisions with certainty. (We will deal with uncertainty in Units 12 and 13.)

Together these three statements fulfill the first four fundamental decision making steps described in Unit 1. Statement 1 specifies the setting and objective, statement 2 the firm's possible decision alternatives, and statement 3 (along with some specific quantitative information supplied shortly) the link between actions and the ultimate objective, namely, profit. It remains for the firm's manager to "solve" and explore this decision problem using marginal analysis (steps 5 and 6).

Before turning to this task, note the simplifying facts embodied in statement

1. Typically, a given firm produces a variety of goods or services. Nonetheless, even for the multiproduct firm, examining products one at a time has significant decision advantages. For one thing, it constitutes an efficient managerial division of labor. Thus, multiproduct firms, such as Procter & Gamble, assign product managers to specific consumer products. A product manager is responsible for charting the future of the brand (pricing, advertising, promotion, and production policies). Similarly, most large companies make profit-maximizing decisions along product lines. This product-by-product strategy is feasible and appropriate as long as the revenues and costs of the firm's products are independent of one another. (As we shall see in Units 3 and 6, things become more complicated if actions taken with respect to one product affect the revenues or costs, or both, of the firm's other products.) In short, the firm can maximize its total profit by separately maximizing the profit derived from each of its product lines.

MARGINAL ANALYSIS

Consider the problem of finding the output level that will maximize the firm's profit. One approach is to use the preceding profit formula and solve the problem by enumeration, that is, by calculating the profits associated with a range of outputs and identifying the one with the greatest profit. Enumeration is a viable approach if there are only a few output levels to test. However, when the number of options is large, enumeration (and the numerous calculations it

requires) is not practical. Instead, we will use the method of marginal analysis to find the "optimal" output level.

Marginal analysis looks at the change in profit that results from making a small change in a decision variable. To illustrate, suppose the firm first considers producing 3 lots, forecasting its resulting profit to be $116,000 as in 2.5. Could it do better than this? To answer this question, the firm considers increasing production slightly, to, say, 3.1 lots. (One-tenth of a lot qualifies as a "small" change. The exact size of the change does not matter as long as it is small.) By substituting Q _ 3.1 into Equation 2.5, we see that the new profit is $117,000. Thus, profit has increased by $1,000. The rate at which profit has changed is a $1,000 increase per .1 lot increase, or 1,000/.1 _ $10,000 per lot.

Marginal Analysis and Calculus

The key to pinpointing the firm's optimal quantity (i.e., the exact output level at which maximum profit is attained) is to compute marginal profit at any given level of output rather than between two nearby output levels. At a particular output, Q, marginal profit is given by the slope of the tangent line to the profit graph at that output level. 2.6 shows an enlarged profit graph with tangent lines drawn at outputs of 3.1, and 3.3 lots. From viewing the tangents, we draw the following simple conclusions. At 3.1 lots, the tangent is upward sloping. Obviously, marginal profit is positive; that is, raising output by a small amount increases total profit. Conversely, at 3.4 lots, the curve is downward sloping. Here marginal profit is negative, so a small reduction in output (not an increase) would increase total profit.

Finally, at 3.3 lots, the tangent is horizontal; that is, the tangent's slope and marginal profit are zero. Maximum profit is attained at precisely this level of output. Indeed, the condition that marginal profit is zero marks this point as the optimal level of output. Remember: If M_ were positive or negative, total profit could be raised by appropriately increasing or decreasing output. Only when M_ is exactly zero have all profit-augmenting opportunities been exhausted. In short, when the profit function's slope just becomes zero, we know we are at the precise peak of the profit curve.3 Thus, we have demonstrated a basic optimization rule:

MARGINAL REVENUE AND MARGINAL COST

The concept of marginal profit yields two key dividends. The general concept instructs the manager that optimal decisions are found by making small changes in decisions, observing the resulting effect on profit, and always moving in the direction of greater profit. A second virtue of the approach is that it provides an efficient tool for calculating the firm's optimal decision. The discussion in this section underscores a third virtue: Marginal analysis is a powerful way to identify the factors that determine profits and, more important, profit changes. We will look once again at the two components of profit, revenue and cost, and highlight the key features of marginal revenue and marginal cost.

Marginal Revenue

Marginal revenue is the amount of additional revenue that comes with a unit increase in output and sales. The marginal revenue (MR) of an increase in unit sales from Q_0 to Q_1 is For instance, the MR earned by increasing sales from 2.0 to 2.1 lots is where 268.8 is the revenue from selling 2.1 lots and 260.0 is the revenue from selling 2.0 lots. The graphic depiction of the MR between two quantities is given by the slope of the line segment joining the two points on the revenue graph. In turn, marginal revenue at a given sales quantity has as its graphic counterpart the slope of the tangent line touching the revenue graph. To calculate the marginal revenue at a given sales output, we start with the revenue expression (Equation 2.3), R _170Q _20Q_2, and take the derivative with respect to quantity: We can use this formula to compute MR at any particular sales quantity. For example, marginal revenue at Q _ 3 is MR _ 170 _ (40)(3) _ $50 thousand; that is, at this sales quantity, a small increase in sales increases revenue at the rate of $50,000 per additional lot sold.

Marginal Cost

Marginal cost (MC) is the additional cost of producing an extra unit of output. The algebraic definition is The computation of MC is particularly easy for the microchip manufacturer's cost function in Equation 2.4. From the cost equation, C _ 100 _ 38Q, it is apparent that producing an extra lot (increasing Q by a unit) will increase cost by $38 thousand. Thus, marginal cost is simply $38 thousand per lot. Note that regardless of how large or small the level of output, marginal cost is always constant. The cost function in Equation 2.4 has a constant slope and thus also an unchanging marginal cost. (We can directly confirm the MC result by taking the derivative of the cost equation.)

SENSITIVITY ANALYSIS

As we saw in Unit 1, sensitivity analysis addresses the basic question: How should the decision maker alter his or her course of action in light of changes in economic conditions? Marginal analysis offers a powerful answer to this question: For any change in economic conditions, we can trace the impact (if any) on the firm's marginal revenue or marginal cost. Once we have identified this impact, we can appeal to the MR _ MC rule to determine the new, optimal decision. 2.9 illustrate the application of this rule for the microchip firm's basic problem. Consider part (a). As before, the firm's decision variable, its output quantity, is listed on the horizontal axis. In turn, levels of MR and MC are shown on the vertical axis, and the respective curves have been graphed. How do we explain the shapes of these curves? For MC, the answer is easy. The marginal cost of producing an extra lot of chips is $38,000 regardless of the starting output level. Thus, the MC line is horizontal, fixed at a level of $38,000. In turn, the graph of the MR curve from Equation 2.8 is

INCREASED MATERIAL COSTS Silicon is the main raw material from which microchips are made. Suppose an increase in the price of silicon causes the firm's estimated cost per lot to rise from $38,000 to $46,000. How should the firm respond? Once again the answer depends on an appeal to marginal analysis. In this case, the firm's MC per chip has changed. In 2.9b, the new MC line lies above and parallel to the old MC line. The intersection of MR and MC occurs at a lower level of output. Because producing extra output has become more expensive, the firm's optimal response is to cut back the level of production. What is the new optimal output? Setting MR _ MC, we obtain 170 _ 40Q _ 46, so Q _ 3.1 lots. In turn, the market-clearing price (using Equation 2.2) is found to be $108,000. The increase in cost has been partially passed on to buyers via a higher price.

INCREASED DEMAND Suppose demand for the firm's chips increases dramatically. At the higher demand, the firm could raise its price by $20,000 per lot ($200 per chip) and still sell the same quantity of chips as before. The old price equation was P _ 170 _ 20Q. The new price equation is P _ 190 _ 20Q. What should be the firm's response? Here the increased demand raises the marginal revenue the firm obtains from selling extra chips. In fact, given the new price equation, the new MR equation must be MR _ 190 _ 40Q. Thus, the new MR curve in 2.9c has a larger intercept than the old one, although the slope is the same. The upward, parallel shift in the MR curve means the new intersection of MR and MC occurs at a higher output. What is the new optimal output? Setting MR _ MC, we find that 190 _ 40Q _ 38, so Q _ 3.8 lots. The corresponding market-clearing price (using the new price equation) is $114,000. The firm takes optimal advantage of the increase in demand by selling a larger output (380 chips per week) at a higher price per lot.

REVIEW QUESTION

1. A manager makes the statement that output should be expanded as long as average revenue exceeds average cost. Does this strategy make sense? Explain.
2. Because of changing demographics, a small, private liberal arts college predicts a fall in enrollments over the next five years. How would it apply marginal analysis to plan for the decreased enrollment? (The college is a non profit institution, so think broadly about its objectives.)
1. Describe A Simple Model of the Firm
2. What is Marginal Analysis?
3. Differentiate about Marginal Revenue and Marginal Cost
4. What is Sensitivity Analysis? Explain.

FURTHER READINGS

1. Managerial Economics - Arun Kumar, Rachana Sharma
2. Managerial Economics - Thomas J. Webster
3. Managerial Economics - Petersen / Jain
4. Managerial Economics - Yogesh Maheshwari
5. Managerial Economics - E. Narayanan Nadar, S. Vijayan

UNIT-3 DEMAND ANALYSIS AND OPTIMAL PRICING

CONTENTS

The Demand Function

To illustrate the basic quantitative aspects of demand, let's start with a concrete example: the demand for air travel.2 Put yourself in the position of a manager for a leading regional airline. One of your specific responsibilities is to analyze the state of travel demand for a nonstop route between Houston, Texas, and a rapidly growing city in Florida. Your airline flies one daily departure from each city to the other (two flights in all) and faces a single competitor that offers two daily flights from each city. Your task is complicated by the fact that the number of travelers on your airline (and therefore the revenue your company earns) has fluctuated considerably in the past three years. Reviewing this past experience, you realize the main determinants of your airline's traffic are your own price and the price of your competitor. In addition, traffic between the two cities was brisk during years in which the Texas and Florida economies enjoyed rapid expansion. But, during the slowdown of 2008, air travel fell between the two cities.

The Demand Curve and Shifting Demand

Suppose that, in the immediate future, regional income is expected to remain at 105 and the competitor's fare will stay at $240. However, your airline's fare is not set in stone, and you naturally are interested in testing the effect of different possible coach prices. Substituting the values of Y and P_ into Equation 3.2's demand function, we find that [3.4] Like the basic demand equation facing the microchip producer in Unit 2, Equation 3.4 relates the quantity of the good or service sold to its price. Here,

however, it is important to remember that, in the background, all other factors affecting demand are held constant (at the values Y _ 105 and P_ _ 240). Of course, it is a simple matter to graph this demand equation as a demand curve.

(Do this yourself as practice.) As usual, the demand curve is downward sloping.5

Starting from an initial price, by varying the coach fare up or down, we move along (respectively up and down) the demand curve. A higher price means lower sales. But what happens if there is a change in one of the other factors that affect demand? As we now show, such a change causes a shift in the demand curve. To illustrate, suppose that a year from now P_ is expected to be unchanged but Y is forecast to grow to 119. What will the demand curve look like a year hence? To answer this question, we substitute the new value, Y _ 119 (along with P_ _ 240), into the demand function to obtain [3.5]

Now compare the new and old demand equations. Observe that they are of the same form, with one key difference: The constant term of the new demand curve is larger than that of the old. Therefore, if your airline were to leave its own fare unchanged a year from now, you would enjoy a greater volume of coach traffic. 3.1 underscores this point by graphing both the old and new demand curves. Note that the new demand curve constitutes a parallel shift to the right (toward greater sales quantities) of the old demand curve. At P _ $240, current demand is 100 seats per flight. At the same fare, coach demand one year from now is forecast to be 142 seats (due to the increase in regional income), a gain of 42 seats. In fact, for any fare your airline might set (and leave unchanged), demand a year from now is predicted to grow by 42 seats. Thus, we confirm that there is a 42-unit rightward shift in the demand curve from old to new demand.

General Determinants of Demand

The example of demand for air travel is representative of the results found for most goods or services. Obviously, the good's own price is a key determinant of demand. (We will say much more about price later in the Unit.) Close behind in importance is the level of income of the potential purchasers of the good or service. A basic definition is useful in describing the effect of income on sales: A product is called a **normal good** if an increase in income raises its sales. In our example, air travel is a normal good. For any normal good, sales vary directly with income; that is, the coefficient on income in the demand equation is positive. As an empirical matter, most goods and services are normal. Any increase in consumer income is spread over a wide variety of goods and services. (Of course, the extra spending on a given good may be small or even nearly zero.) Likewise, when income is reduced in an economy that is experiencing a recession, demand falls across the spectrum of normal goods. For a small category of goods (such as certain food staples), an increase in income causes a reduction in spending. These are termed **inferior goods.** For instance, an individual of moderate means may regularly consume a large quantity of beans, rice, and ground meat. But, after experiencing an increase in income, the individual can better afford other foods and therefore reduces his consumption of the old staples.

A third set of factors affecting demand are the prices of substitute and complementary goods. As the term suggests, a **substitute good** competes with and can substitute for the good in question. In the airline example, travel on one airline serving the same intercity route is a very close substitute for travel on the other. Accordingly, an increase in the price of the substitute good or service causes an increase in demand for the good in question (by making it relatively more attractive to purchase). Note that substitution in demand can occur at many levels. For instance, the airline's sales along the route are affected not only by changes in competing airline fares but also by train and bus fares and auto-operating costs. To a greater or lesser degree, these other modes of transportation are substitutes for air travel.

A pair of goods is **complementary** if an increase in demand for one causes an increase in demand for the other. For instance, an increase in the sales of new automobiles will have a positive effect on the sales of new tires. In particular, tire manufacturers are very interested in the prices car manufacturers announce for new models. They know that discount auto prices will spur not only the sales of cars, but also the sales of tires. The price of a complementary

good enters negatively into the demand function; that is, an increase in the price of a complementary good reduces demand for the good in question. For example, Florida resort packages and travel between Houston and Florida are to some extent complementary. Thus, the price of resort packages would enter with a negative coefficient into the demand function for travel along the route.6

Finally, a wide variety of other factors may affect the demand for particular goods and services. Normal population growth of prime groups that consume the good or service will increase demand. As the populations of Houston and the Florida city grow, so will air travel between them. The main determinant of soft-drink sales is the number of individuals in the 10-to-25 age group. Changes in preferences and tastes are another important factor. Various trends over the past 20 years have supported growth in demand for new foods (diet, natural, organic), new electronic products (cell phones, digital cameras, MP3 players, CD and DVD players), new recreation services (exercise, travel, tanning salons, and so on). The list is endless.

ELASTICITY OF DEMAND

Price Elasticity

Price elasticity measures the responsiveness of a good's sales to changes in its price. This concept is important for two reasons. First, knowledge of a good's price elasticity allows firms to predict the impact of price changes on unit sales. Second, price elasticity guides the firm's profit-maximizing pricing decisions.

Let's begin with a basic definition: The **price elasticity of demand** is the ratio of the percentage change in quantity and the percentage change in the good's price, all other factors held constant. In algebraic terms, we have where P_0 and Q_0 are the initial price and quantity, respectively.

Price elasticity is a key ingredient in applying marginal analysis to determine optimal prices. Because marginal analysis works by evaluating "small" changes taken with respect to an initial decision, it is useful to

measure elasticity with respect to an infinitesimally small change in price. In this instance, we write elasticity as

$$E_p = \frac{\%\text{ change in Q}}{\%\text{ change in P}} = \frac{\Delta Q/Q}{\Delta P/P} = \frac{(Q_1 - Q_0)/Q_0}{(P_1 - P_0)/P_0}$$

$$E_p = \frac{dQ/Q}{dP/P}.$$

We can rearrange this expression to read

$$E_p = \left(\frac{dQ}{dP}\right)\left(\frac{P}{Q}\right).$$

The algebraic expressions in Equations 3.7 and 3.8a are referred to as point elasticities because they link percentage quantity and price changes at a pricequantity point on the demand curve. Although most widely used, point elasticity measures are not the only way to describe changes in price and quantity.

The main advantage of the arc elasticity measure is that it treats the prices and quantities symmetrically; that is, it does not distinguish between the "initial" and "final" prices and quantities. Regardless of the starting point, the elasticity is the same. In contrast, in computing the elasticity via Equation 3.7, one must be careful to specify P0 and Q0. To illustrate, suppose the initial airfare is $235 and 110 seats are filled. The elasticity associated with a price hike to $240 (and a drop to 100 seats) is EP _ (_10/110)/(5/235) __4.3. Thus, we see that the elasticity associated with the change is _4.8 or _4.3, depending on the starting point. The overriding advantage of point elasticities (Equation 3.8a) is their application

in conjunction with marginal analysis. For instance, a firm's optimal pricing policy depends directly on its estimate of the price elasticity, EP _ (dQ/Q)/(dP/P). In this and later Units, we will focus on point elasticities in our analysis of optimal decisions.7

Elasticity measures the sensitivity of demand with respect to price. In describing elasticities, it is useful to start with a basic benchmark. First, demand is said to be **unitary elastic** if EP _ _1. In this case, the percentage change in price is exactly matched by the resulting percentage change in quantity, but in the opposite direction. Second, demand is **inelastic** if _1 _ EP _ 0. The term inelastic suggests that demand is relatively unresponsive to price: The percentage change in quantity is less (in absolute value) than the percentage change in price. Finally, demand is **elastic** if EP__1. In this case, an initial change in price causes a larger percentage change in quantity. In short, elastic demand is highly responsive, or sensitive, to changes in price.

The easiest way to understand the meaning of inelastic and elastic demand is to examine two extreme cases. 3.2a depicts a vertical demand

curve representing **perfectly inelastic** demand, $E_P = 0$. Here sales are constant (at $P = (P_0 + P_1)/2$.

FACTORS AFFECTING PRICE ELASTICITY What determines whether the demand for a good is price elastic or price inelastic? Here are four important factors.

A first factor is the degree to which the good is a necessity. If a good or service is not considered essential, the purchaser can easily do without it—if and when the price becomes too high—even if there are no close substitutes. In that case, demand is elastic. If the good is a necessary component of consumption, it is more difficult to do without it in the face of a price increase. Thus, demand tends to be price inelastic.

A second factor is the availability of substitutes. With many substitutes, consumers easily can shift to other alternatives if the price of one good becomes too high; demand is elastic. Without close substitutes, switching becomes more difficult; demand is more inelastic. For this reason, industry demand tends to be much less elastic than the demand facing a particular firm in the industry. If one firm's price increases, consumers are able to go to other firms quite easily. Thus, the demand facing a single firm in an industry may be quite elastic because competitors produce goods that are close substitutes. But consider what happens if the industry price goes up, that is, all firms in the industry increase their prices in unison. In this case, price-sensitive consumers are limited in their course of action: to do without the good or to find a good in another industry to replace it. If these options are infeasible, the third option is to pay the higher price. Thus, industry demand is less elastic. The same point applies to the case where a single monopolist dominates an industry or product line. Other things being equal, the monopolist's demand is less elastic (since it is the sole producer) than the demand facing a particular firm in a multifirm industry. A third determinant of price elasticity is the proportion of income a consumer spends on the good in question. The issue here is the cost of searching for suitable alternatives to the good. It takes time and money to compare substitute products. If an individual spends a significant portion of income on a good, he or she will find it worthwhile to search for and compare the prices of other goods. Thus, the consumer is price sensitive. If spending on the good represents only a small portion of total income, however, the search for substitutes will not be worth the time, effort, and expense. Thus, other things being equal, the demand for small-ticket items tends to be relatively inelastic. Finally, time of adjustment is an important influence on elasticity. When the price of gasoline dramatically increased in the last five years, consumers initially had little recourse but to pay higher prices at the pump. Much of the population continued to drive to work in large, gas-guzzling cars. As time passed, however, consumers began to make adjustments. Some commuters have now switched from automobiles to buses or other means of public transit. Gas guzzlers have been replaced by smaller, more fuel-efficient cars including hybrids. Some workers have moved closer to their jobs, and when jobs turn over, workers have found new jobs closer to their homes. Thus, in the short run, the demand for gasoline is relatively inelastic. But in the long run, demand appears to be much more elastic as people are able to cut back

consumption by a surprising amount. Thus, the time of adjustment is crucial. As a general rule, demand is more elastic in the long run than in the short run.

Other Elasticities

The elasticity concept can be applied to any explanatory variable that affects sales. Many of these variables—income, the prices of substitutes and complements, and changes in population or preferences—have already been mentioned. (An additional important variable affecting sales is the firm's spending on advertising and promotion.) To illustrate, consider the elasticity of demand with respect to income (Y). This is defined as in a manner exactly analogous to the earlier price elasticity definition.9 **Income elasticity** links percentage changes in sales to changes in income, all other

$$E_Y = \frac{\%\text{ change in Q}}{\%\text{ change in Y}} = \frac{\Delta Q/Q}{\Delta Y/Y}$$

A main impact on the sales outlook for an industry, a firm, or a particular good or service is the overall strength of the economy. When the economy grows strongly, so do personal income, business profits, and government income. Gains in these income categories generate increased spending on a wide variety of goods and services. Conversely, when income falls during a recession, so do sales across the economy. Income elasticity thus provides an important measure of the sensitivity of sales for a given product to swings in the economy. For instance, if EY _ 1, sales move exactly in step with changes in income. If EY _ 1, sales are highly cyclical, that is, sensitive to income. For an inferior good, sales are countercyclical, that is, move in the opposite direction of income and EY _ 0.

CROSS-PRICE

Price Elasticity and Prediction

Price elasticity is an essential tool for estimating the sales response to possible price changes. A simple rearrangement of the elasticity definition (Equation 3.7) gives the predictive equation

$$\Delta Q/Q = E_P(\Delta P/P)$$

For instance, in Table 3.1, the short-term (i.e., one-year) price elasticity of demand for gasoline is approximately _.3. This indicates that if the average price of gasoline were to increase from $2.50 to $3.00 per gallon (a 20 percent increase), then consumption of gasoline (in gallons) would fall by only 6 percent (_.3 _ 20%). The table also shows that the price elasticity of demand for luxury cars is _2.1. A modest 5 percent increase in their average sticker price implies a 10.5 percent drop in sales. (Caution: Equation 3.9 is exact for very small changes but only an approximation for large percentage changes, over which elasticities may vary.)

$$\Delta Q/Q = E_P(\Delta P/P) + E_Y(\Delta Y/Y)$$

DEMAND ANALYSIS AND OPTIMAL PRICING

In this section, we put demand analysis to work by examining three important managerial decisions: (1) the special case of revenue maximization, (2) optimal markup pricing, and (3) price discrimination.

Price Elasticity, Revenue, and Marginal Revenue

What can we say about the elasticity along any downward-sloping, linear demand curve? First, we must be careful to specify the starting quantity and price (the point on the demand curve) from which percentage changes are measured. From Equation 3.8b, we know that EP _ (dQ/dP)(P/Q). The slope of the demand curve is dP/dQ (as it is conventionally drawn with price on the vertical axis). Thus, the first term in the elasticity expression, dQ/dP, is simply the inverse of this slope and is constant everywhere along the curve. The term P/Q decreases as one moves downward along the curve. Thus, along a linear demand curve, moving to lower prices and greater quantities reduces elasticity; that is, demand becomes more inelastic.

As a concrete illustration of this point, consider a software firm that is trying to determine the optimal price for one of its popular software programs. Management estimates this product's demand curve to be where Q is copies sold per week and P is in dollars. We note for future reference that dQ/dP _ _4. 3.3a shows this demand curve as well as the associated marginal revenue curve. In the , the midpoint of the demand curve is marked by point M: Q _ 800 and P _ $200. Two other points, A and B, along the demand curve also are shown.

$$Q = 1{,}600 - 4P,$$

Maximizing Revenue

As we saw in Unit 2, there generally is a conflict between the goals of maximizing revenue and maximizing profit. Clearly, maximizing profit is the appropriate objective because it takes into account not only revenues but also relevant costs. In some important special cases, however, the two goals coincide or are equivalent. This occurs when the firm faces what is sometimes called a **pure selling problem:** a situation where it supplies a good or service while incurring no variable cost (or a variable cost so small that it safely can be ignored). It should be clear that, without any variable costs, the firm maximizes its ultimate profit by setting price and output to gain as much revenue as possible (from which any fixed costs then are paid). The following pricing problems serve as examples.

- A software firm is deciding the optimal selling price for its software.
- A manufacturer must sell (or otherwise dispose of) an inventory of unsold merchandise.
- A professional sports franchise must set its ticket prices for its home games.
- An airline is attempting to fill its empty seats on a regularly scheduled flight.

In each of these examples, variable costs are absent (or very small). The cost of an additional software copy (documentation and disk included) is trivial. In the case of airline or sports tickets, revenues crucially depend on how many tickets are sold. The cost of an additional passenger or

spectator is negligible once the flight or event has been scheduled. As for inventory, production costs are sunk; selling costs are negligible or very small. Thus, in each case the firm maximizes profits by setting price and output to maximize revenue. How does the firm determine its revenue-maximizing price and output? There are two equivalent answers to this question. The first answer is to apply Unit 2's fundamental rule: MR _ MC. In the case of a pure selling problem, marginal cost is zero. Thus, the rule becomes MR _ 0, exactly as one would expect. This rule instructs the manager to push sales to the point where there is no more additional revenue to be had—MR _ 0—and no further. From the preceding discussion, we have established a second, equivalent answer: Revenue is maximized at the point of unitary elasticity. If demand were inelastic or elastic, revenue could be increased by raising or lowering price, respectively. The following proposition sums up these results. Revenue is maximized at the price and quantity for which marginal revenue is zero or, equivalently, the price elasticity of demand is unity (_1). Note that this result confirms that the point of unitary elasticity occurs at the midpoint of a linear demand curve. For the sales quantity at the midpoint, marginal revenue is exactly zero (since the MR curve cuts the horizontal axis at the midpoint quantity). But when MR _ 0, it is also true that EP__1.

Optimal Markup Pricing

There is a close link between demand for a firm's product and the firm's optimal pricing policy. In the remainder of this Unit, we will take a close and careful look at the trade-off between price and profit. Recall that in Unit 2, the focus was squarely on the firm's quantity decision. Once the firm determined its optimal output by weighing marginal revenue and marginal cost, it was a simple matter to set price in order to sell exactly that much output. Now we shift our focus to price and consider a somewhat different trade-off.

Business Behavior: Pricing in Practice

Our study of optimal managerial decisions suggests two points of criticism about full-cost pricing. First, full-cost pricing uses average cost—the incorrect measure of relevant cost—as its base. The logic of marginal analysis in general and the optimal markup rule (Equation 3.13) in particular show that optimal price and quantity depend on marginal cost. Fixed costs, which are counted in AC but not in MC, have no effect on the choice of optimal price and quantity.[12] Thus, to the extent that AC differs from MC, the full-cost method can lead to pricing errors. Second, the percentage markup should depend on the elasticity of demand. There is considerable evidence that firms vary their markups in rough accord with price elasticity.[13] Gourmet frozen foods carry much higher markups than generic food items. Inexpensive digital watches ($15 and under) have lower markups than fine Swiss watches or jewelers' watches. Designer dresses and wedding dresses carry much higher markups than off-the-rack dresses. In short, producers' markups are linked to elasticities, at least in a qualitative sense. Nonetheless, it is unlikely that firms' full-cost markups exactly duplicate optimal markups. Obviously, a firm that sets a fixed markup irrespective of elasticity is needlessly sacrificing profit.

Price Discrimination

Price discrimination occurs when a firm sells the same good or service to different buyers at different prices.[15] As the following examples suggest, price discrimination is a common business practice.

• Airlines charge full fares to business travelers, while offering discount fares to vacationers.

• Firms sell the same products under different brand names or labels at different prices.

• Providers of professional services (doctors, consultants, lawyers, etc.) set different rates for different clients.

• Manufacturers introduce products at high prices before gradually dropping price over time.

• Publishers of academic journals charge much higher subscription rates to libraries and institutions than to individual subscribers.

• Businesses offer student and senior citizen discounts for many goods and services.

• Manufacturers sell the same products at higher prices in the retail market than in the wholesale market.

• Movies play in "first-run" theaters at higher ticket prices before being released to suburban theaters at lower prices.

When a firm practices price discrimination, it sets different prices for different market segments, even though its costs of serving each customer group are the same. Thus, price discrimination is purely demand based. Of course, firms may also charge different prices for the "same" good or service because of cost differences. (For instance, transportation cost may be one reason why the same make and model of automobile sells for significantly different prices on the West and East coasts.) But cost-based pricing does not fall under the heading of price discrimination.

Price discrimination is a departure from the pricing model we have examined up to this point. Thus far, the firm has been presumed to set a single market-clearing price. Obviously, charging different prices to different market segments, as in the examples just listed, allows the firm considerably more pricing flexibility. More to the point, the firm can increase its profit with a policy of optimal price discrimination (when the opportunity exists).

Multinational Production and Pricing Revisited

In the first example in Unit 1, an automobile producer faced the problem of pricing its output at home and abroad. We are now ready to put demand analysis to work to determine the firm's optimal decisions. The facts are as follows: The producer faces relatively little competition at home; it is one of the most efficient domestic producers, and trade barriers limit the import of foreign cars. However, it competes in the foreign market with many local and foreign manufacturers. Under these circumstances, demand at home is likely to be much more inelastic than demand in the foreign country. Suppose that the price equations at home (H) and abroad (F) are, respectively,

DEMAND-BASED PRICING

As these examples indicate, the ways in which firms price discriminate are varied. Indeed, there are many forms of demand-based P_H _ 30,000 _ $50Q_H$ and P_F _ 25,000 _ $70Q_F$ pricing that are closely related to price discrimination (although not always called by that name). For instance, resorts in Florida and the Caribbean set much higher nightly rates during the high season (December to March) than at off-peak times. The difference in rates is demand based. (The resorts' operating costs differ little by season.) Vacationers are willing to pay a much higher price for warm climes during the North American winter. Similarly, a convenience store, open 24 hours a day and located along a high-traffic route or intersection, will set premium prices for its merchandise. (Again, the high markups are predominantly demand based and only partly based on higher costs.) Likewise, golf courses charge much higher prices on weekends than on weekdays. Each of these examples illustrates demand-based pricing.

FORMS OF PRICE DISCRIMINATION

It is useful to distinguish three forms of price discrimination. The practice of charging different prices to different market segments (for which the firm's costs are identical) is often referred to as **third-degree price discrimination.** Airline and movie ticket pricing are examples. Prices differ across market segments, but customers within a market segment pay the same price.

Now suppose the firm could distinguish among different consumers within a market segment. What if the firm knew each customer's demand curve? Then it could practice perfect price discrimination. **First-degree**, or **perfect, price discrimination** occurs when a firm sets a different price for each customer and by doing so extracts the maximum possible sales revenue. As an example, consider an auto dealer who has a large stock of used cars for sale and expects 10 serious potential buyers to enter her showroom each week. She posts different model prices, but she knows (and customers know) that the sticker price is a starting point in subsequent negotiations. Each customer knows the maximum price he or she is personally willing to pay for the car in question. If the dealer is a shrewd judge of character, she can guess the range of each buyer's maximum price and, via the negotiations, extract almost this full value. For instance, if four buyers' maximum prices are $6,100, $6,450, $5,950, and $6,200, the perfectly discriminating dealer will negotiate prices nearly equal to these values. In this way, the dealer will sell the four cars for the maximum possible revenue. As this example illustrates, perfect discrimination is fine in principle but much more difficult in practice. Clearly, such discrimination requires that the seller have an unrealistic amount of information. Thus, it serves mainly as a benchmark— a limiting case at best.

Finally, **second-degree price discrimination** occurs when the firm offers different price schedules, and customers choose the terms that best fit their needs. The most common example is the offer of quantity discounts: For large volumes, the seller charges a lower price per unit, so the buyer purchases a larger quantity. With a little thought, one readily recognizes this as a form of profitable price discrimination. High-

volume, price-sensitive buyers will choose to purchase larger quantities at a lower unit price, whereas low-volume users will purchase fewer units at a higher unit price. Perhaps the most common form of quantity discounts is the practice of two-part pricing. As the term suggests, the total price paid by a customer is where A is a fixed fee (paid irrespective of quantity) and p is the additional price per unit. Telephone service, electricity, and residential gas all carry twopart prices. Taxi service, photocopy rental agreements, and amusement park admissions are other examples. Notice that two-part pricing implies a quantity discount; the average price per unit, P/Q _ A/Q _ p, declines as Q increases. Two-part pricing allows the firm to charge customers for access to valuable services (via A) while promoting volume purchases (via low p).

Information Goods

In the last 20 years, we have witnessed explosive growth in the provision of **information goods and services**. The business press speaks of Internet industries and e-business markets. The "information" label is meant to be both more broad based and more precise. An information good could be a database, game cartridge, news article (in electronic or paper form), piece of music, or piece of software. Information services range from e-mail and instant messaging to electronic exchanges and auctions, to brokerage and other financial services, to job placements. Of course, information services also include all manner of Internet-based transactions, such as purchasing airline tickets, selling real estate, procuring industrial inputs, and gathering extensive data on potential customers.

CUSTOMIZED PRICING AND PRODUCTS.

The emergence of electronic commerce and online transactions has greatly expanded the opportunities for market segmentation and price discrimination. From management's point of view, the beauty of information goods and services is that they can be sold over and over again (at negligible marginal cost). Moreover, unlike a traditional good sold at a posted price from a store shelf, the price of an information good (transacted electronically) can be changed minute by minute, customer by customer. Sellers of sophisticated databases—from Reuters to Lexis-Nexis to Bloomberg financial information—set scores of different prices to different customers. As always, prices are set according to elasticities; the most price-sensitive (elastic) customers receive the steepest discounted prices. Consider the ways in which an airline Web site (such as www.delta.com) can price its airline seats. Each time a customer enters a possible itinerary with departure and return dates, the Web page responds with possible flights and prices. These electronic prices already reflect many features: the class of seat, 21-day, 14-day, or 7-day advanced booking, whether a Saturday night stay is included, and so on. By booking in advance and staying a Saturday night, pleasure travelers can take advantage of discounted fares. Business travelers, whose itineraries are not able to meet these restrictions, pay much higher prices. Moreover, the airline can modify prices instantly to reflect changes in demand. If there is a surplus of unsold discount seats as the departure date approaches, the airline can further cut their price or sell the seats as part of a vacation package (hotel stay, rentacar included) at even a

steeper discount. (Airlines also release seats to discount sellers, such as Priceline.com, Hotwire.com, and lastminute.com, who sell tickets at steep discounts to the most price-sensitive fliers.) Or some discount seats might be reassigned as full-fare seats if last-minute business demand for the flight is particularly brisk. Online, the pricing possibilities are endless.

Consumer Preferences and Demand

In this appendix, we provide a brief overview of the foundations of consumer demand—how consumers allocate their spending among desired goods and services. The analysis is important in its own right as a basis for downward sloping demand curves. Perhaps its greater importance lies in the broader decision-making principle it illustrates. As we shall see, an optimal decision— made either by a consumer or a manager—depends on a careful analysis of preferences and trade-offs among available alternatives.

The Consumer's Problem

Consider an individual who must decide how to allocate her spending between desirable goods and services. To keep things simple, let's limit our attention to the case of two goods, X and Y. These goods could be anything from specific items (soft drinks versus bread) to general budget categories (groceries versus restaurant meals or food expenditures versus travel spending). The consumer faces a basic question: Given a limited amount of money to spend on the two goods, and given their prices, what quantities should she purchase?

INDIFFERENCE CURVES

To answer this question, we will use a simple graphical device to describe the individual's preferences. Imagine that we have asked the consumer what her preferences are for alternative bundles of goods. Which do you prefer, 5 units of X and 10 units of Y, or 7 units of X and 6 units of Y? The answers to enough of such questions generate a preference ranking for a wide range of possible bundles of goods. 3A.1 shows these possible bundles by listing the quantities of the goods on the respective axes. The also depicts a number of the consumer's indifference curves as a way of representing her preferences.

As its name suggests, an **indifference curve** shows all combinations of the goods among which the individual is indifferent. The consumer is indifferent between all bundles on the same curve. Using the middle indifference curve in the , we see that the consumer is indifferent between the bundle containing 15 units of Y and 2 units of X (point A), 10 units of Y and 3 units of X (point B), and 4 units of Y and 6 units of X (point D). The bundles corresponding to points C, E, and F lie on the same indifference curve and are equally preferred by the consumer.

We can make three observations about the consumer's indifference curves. First, as we move to greater quantities of both goods, we move to higher and higher indifference curves. The depicts three different indifference curves. The consumer's welfare increases as we move to curves farther to the northeast in the .1 Second, we note that the indifference curve is downward sloping. Since both goods are valued by the consumer, a decrease in one good must be compensated by an

increase in the other to maintain the same level of welfare (or utility) for the consumer.

Third, we note that the slope of each curve goes from steep to flat, moving southeast along its length. This means that the trade-off between the goods changes as their relative quantities change. For instance, consider a movement from A to B. At point A, the consumer has 15 units of Y (a relative abundance) and 2 units of X. By switching to point B, she is willing to give up 5 units of Y to gain a single additional unit of X. Thus, the trade-off is five to one. By moving from point B (where Y is still relatively abundant) to point C, the consumer is willing to give up another 3 units of Y to get an additional unit of X. Now the trade-off between the goods (while leaving the consumer indifferent) is three to one. The trade-offs between the goods continue to diminish by movements from C to D to E. Thus, the indifference curve is bowed. This shape represents a general result about consumer preferences: The greater the amount of a good a consumer has, the less an additional unit is worth to him or her. This result usually is referred to as the law of diminishing marginal utility. In our example, moving southeast along the indifference curve means going from a relative abundance of Y and a scarcity of X to the opposite proportions. When X is scarce, the consumer is willing to trade many units of Y for an additional unit of X. As X becomes more abundant and Y more scarce, X's relative value diminishes and Y's relative value increases.

THE BUDGET CONSTRAINT

Having described her preferences, next we determine the consumer's alternatives. The amount of goods she can purchase depends on her available income and the goods' prices. Suppose the consumer sets aside $20 each week to spend on the two goods. The price of good X is $4 per unit, and the price of Y is $2 per unit. Then she is able to buy any quantities of the goods (call these quantities X and Y) as long as she does not exceed her income. If she spends the entire $20, her purchases must satisfy [3A.1]

This equation's left side expresses the total amount the consumer spends on the goods. The right side is her available income. According to the equation, her spending just exhausts her available income.[2] This equation is called the consumer's **budget constraint**. 3A.2 depicts the graph of this constraint. For instance, the consumer could purchase 5 units of X and no units of Y (point A), 10 units of Y and no units of X (point C), 3 units of X and 4 units of Y (point B), or any other combination along the budget line shown. Note that bundles of goods to the northeast of the budget line are infeasible; they cost more than the $20 that the consumer has to spend.

OPTIMAL CONSUMPTION

We are now ready to combine the consumer's indifference curves with her budget constraint to determine her optimal purchase quantities of the goods. 3A.3 shows that the consumer's optimal combination of goods lies at point B, 3 units of X and 4 units of Y. Bundle B is optimal precisely because it lies on the consumer's "highest" attainable indifference curve while satisfying the budget constraint. (Check that all other bundles along the budget line lie on lower indifference curves.)

Demand Curves

The demand curve graphs the relationship between a good's price and the quantity demanded, holding all other factors constant. Consider the consumer's purchase of good X as its price is varied (holding income and the price of Y constant). What if the price falls from $4 per unit to $2 per unit to $1 per unit? 3A.4 shows the effect of these price changes on the consumer's budget line. As the price falls from $4 to $2, the budget line flattens and pivots around its vertical intercept. (Note that, with the price of Y unchanged, the maximum amount of Y the consumer can purchase remains the same.) The shows the new budget lines and new points of optimal consumption at the lower prices. As one would expect, reduction in price brings forth greater purchases of good X and increases the consumer's welfare (i.e., she moves to higher

REVIEW QUESTION

1. Define the demand function. What is the demand curve and shifting demand?
2. What are the general determinants of demand?
3. Describe elasticity of demand. What are the factors affecting price elasticity?
4. Describe demand analysis and optimal pricing in detail.
5. What is optimal markup pricing? Explain.
6. What are the causes of price discrimination? Explain forms of price discrimination.
7. Describe customized pricing and products. What is consumer preferences and demand?

FURTHER READINGS

1. Managerial Economics - Arun Kumar, Rachana Sharma
2. Managerial Economics - Thomas J. Webster
3. Managerial Economics - Petersen / Jain
4. Managerial Economics - Yogesh Maheshwari
5. Managerial Economics - E. Narayanan Nadar, S. Vijayan

UNIT-4 ESTIMATING AND FORECASTING DEMAND

CONTENTS

INTRODUCTION

This Unit is organized as follows. We begin by examining sources of information that provide data for forecasts. These include consumer interviews and surveys, controlled market studies, and uncontrolled market data. Next, we explore regression analysis, a statistical method widely used in demand estimation. Finally, we consider a number of important forecasting methods.

COLLECTING DATA

Consumer Surveys

A direct way to gather information is to ask people. Whether face to face, by telephone, online, or via direct mail, researchers can ask current and prospective customers a host of questions: How much of the product do you plan to buy this year? What if the price increased by 10 percent? Do price rebates influence your purchase decisions, and, if so, by how much? What features do you value most? Do you know about the current advertising campaign for the product? Do you purchase competing products? If so, what do you like about them?

Consumer product companies use surveys extensively. In a given year, Campbell Soup Company questions over 100,000 consumers about foods and uses the responses to modify and improve its product offerings and to construct demand equations. Marriott Corporation used this method to design the Courtyard by Marriott hotel chain, asking hundreds of interviewees to compare features and prices. Today, the explosion of online surveys allows firms to collect thousands of responses (often highly detailed) at very low cost.

SURVEY PITFALLS

Though useful, surveys have problems and limitations. For example, market researchers may ask the right questions, but of the wrong people. Economists call this **sample bias**. In some contexts, random sampling protects against sample bias. In other cases, surveys must take care in targeting a representative sample of the relevant market segment.

A second problem is **response bias**. Respondents might report what they believe the questioner wants to hear. ("Your product is terrific, and I intend to buy it this year if at all possible.") Alternatively, the customer may attempt to influence decision making. ("If you raise the price, I definitely will stop buying.") Neither response will likely reflect the potential customer's true preferences. A third problem is **response accuracy**. Even if unbiased and forthright, a potential customer may have difficulty in answering a question accurately. ("I think I might buy it at that price, but when push comes to shove, who knows?") Potential customers often have little idea of how they will react to a price increase or to an increase in advertising. A final difficulty is **cost**.

Conducting extensive consumer surveys is extremely costly. As in any economic decision, the costs of acquiring additional information must be weighed against the benefits.

Controlled Market Studies

Firms can also generate data on product demand by selling their product in several smaller markets while varying key demand determinants, such as price, across the markets. The firm might set a high price with high advertising spending in one market, a high price and low advertising in another, a low price and high advertising in yet another, and so on. By observing sales responses in the different markets, the firm can learn how various pricing and advertising policies (and possible interactions among them) affect demand. To draw valid conclusions from such market studies, all other factors affecting demand should vary as little as possible across the markets. The most common—and important—of these "other" demand factors include population size, consumer incomes and tastes, competitors' prices, and even differences in climate. Unfortunately, regional and cultural differences, built-up brand loyalties, and other subtle but potentially important differences may thwart the search for uniform markets. In practice, researchers seek to identify and control as many of these extraneous factors as possible.

Uncontrolled Market Data

In its everyday operation, the market itself produces a large amount of data. Many firms operate in multiple markets. Population, income, product features, product quality, prices, and advertising vary across markets and over time. All of this change creates both opportunity and difficulty for the market researcher. Change allows researchers to see how changing factors affect demand. With uncontrolled markets, however, many factors change at the same time. How, then, can a firm judge the effect of any single factor? Fortunately, statisticians have developed methods to handle this very problem. During the last 20 years, firms have increasingly used sophisticated computer- based methods to gather market data. Today more than three-quarters of all supermarkets

employ check-out scanners that provide enormous quantities of data about consumer purchases. Internet purchases provide an expanding universe of additional data on consumer preferences and purchasing behavior.

REGRESSION ANALYSIS

Regression analysis is a set of statistical techniques using past observations to find (or estimate) the equation that best summarizes the relationships among key economic variables. The method requires that analysts (1) collect data on the variables in question, (2) specify the form of the equation relating the variables, (3) estimate the equation coefficients, and (4) evaluate the accuracy of the equation. Let's begin with a concrete example.

Interpreting Regression Statistics

Many computer programs are available to carry out regression analysis. (In fact, almost all of the best-selling spreadsheet programs include regression features.) These programs call for the user to specify the form of the regression equation and to input the necessary data to estimate it: values of the dependent variables and the chosen explanatory variables. Besides computing the ordinary least-squares regression coefficients, the program produces a set of statistics indicating how well the OLS equation performs. Table 4.6 lists the standard computer output for the airline's multiple regression. The regression coefficients and constant term are listed in the third-to-last line. Using these, we obtained the regression equation: To evaluate how well this equation fits the data, we must learn how to interpret the other statistics in the table.

R-SQUARED

The **R-squared statistic** (also known as the coefficient of determination) measures the proportion of the variation in the dependent variable (Q in our example) that is explained by the multiple-regression equation. Sometimes we say that it is a measure of goodness of fit, that is, how well the equation fits the data. The total variation in the dependent variable is computed as that is, as the sum across the data set of squared differences between the values of Q and the mean of Q. In our example, this total sum of squares (labeled TSS) happens to be 11,706. The sum of squared errors, SSE, embodies the variation in Q not accounted for by the regression equation. Thus, the numerator is the amount of explained variation and R-squared is simply the ratio of explained to total variation. In our example, we can calculate that $R2$ _ (11,706 _ 2,616)/11,706 _ .78. This confirms the entry in Table 4.6. We can rewrite Equation 4.4 as

$$R^2 = \frac{TSS - SSE}{TSS}$$

Potential Problems in Regression

Regression analysis can be quite powerful. Nonetheless, it is important to be aware of the limitations and potential problems of the regression approach.

EQUATION SPECIFICATION

In our example, we assumed a linear form, and the resulting equation tracked the past data quite well. However, the real world is not always linear; relations do not always follow straight lines. Thus, we may be making an error in specification, and this can lead to poorer predictions. The **constant elasticity** demand equation also is widely used.

OMITTED VARIABLES

A related problem is that of **omitted variables**. Recall that we began the analysis of airline demand with price as the only explanatory variable. The resulting OLS equation produced predictions that did a reasonably good job of tracking actual values. However, a more comprehensive equation, accounting for competitor's price and income, did far better. In short, leaving out key variables necessarily worsens prediction performance. In fact, omission of these other variables also affects the coefficients of the included variables. For instance, the price coefficient is _1.63 when it is the sole explanatory variable. This is quite different from the estimated multiple regression coefficient, _2.12. Thus, the single-variable regression underestimates the magnitude of the true price effect.

MULTICOLLINEARITY

When two or more explanatory variables move together, we say that the regression suffers from **multi collinearity**. In this case, it is difficult to tell which of the variables is affecting the dependent variable. Suppose demand for a firm's product is believed to depend on only two factors: price and advertising. The data show that whenever the firm initiated an aggressive advertising campaign, it invariably lowered the good's price. Sales increased significantly as a result. When the firm decreased advertising spending it also increased price, and sales dropped. The question is: Should the changes in sales be attributed to changes in advertising or to changes in price? Unfortunately, it is impossible to tell, even with regression. If two right-hand variables move together, regression cannot separate the effects. Regression does not require that we hold one of the factors constant as we vary the other, but it does require that the two factors vary in different ways.

What happens when the forecaster runs a regression based on these data? If the right-hand variables are perfectly correlated, the computerized regression program will send back an error message. If the right-hand variables are not perfectly correlated, but move very closely together (either directly or inversely), the regression output will provide very imprecise coefficient estimates with large standard errors. In this case, additional data may improve the estimates. If not, the forecaster must live with the imprecise estimates.

Can the firm still use the equation to forecast? Yes and no. It can if it plans to continue the pattern of lowering price whenever it increases advertising. In that case, it need not care about the separate effects. However, if it plans to lower price without an advertising campaign, or to advertise more without lowering price, the forecast will be very unreliable.

SIMULTANEITY AND IDENTIFICATION

This brings us to a subtle, but interesting and important, issue. In the preceding discussion, we assumed that the firm had explicit control over its price. In many settings, however, price is determined by overall demand and supply conditions, not by the individual firm. Here, the firm must take the price the market dictates or else sell nothing.

FORECASTING

Forecasting models often are divided into two main categories: structural and nonstructural models. Structural models identify how a particular variable of interest depends on other economic variables. The airline demand equation (4.3) is a single- equation structural model. Sophisticated large-scale structural models of the economy often contain hundreds of equations and more than a thousand variables and usually are referred to as econometric models. Nonstructural models focus on identifying patterns in the movements of economic variables over time. One of the best-known methods, time-series analysis, attempts to describe these patterns explicitly. A second method, barometric analysis, seeks to identify leading indicators—economic variables that signal 150 Unit 4 Estimating and Forecasting Demand future economic developments. (The stock market is one of the best-known leading indicators of the course of the economy.)

Time-Series Models

Time-series models seek to predict outcomes simply by extrapolating past behavior into the future. Time-series patterns can be broken down into the following four categories.

1. Trends
2. Business cycles
3. Seasonal variations
4. Random fluctuations

A **trend** is a steady movement in an economic variable over time. For example, the total production of goods and services in the United States (and most other countries) has moved steadily upward over the years. Conversely, the number of farmers in the United States has steadily declined. On top of such trends are periodic **business cycles**. Economies experience periods of expansion marked by rapid growth in gross domestic product (GDP), investment, and employment. Then economic growth may slow and even fall. A sustained fall in (real) GDP and employment is called a recession. For the United States' economy, recessions have become less frequent and less severe since 1945. Nonetheless, the business cycle—with periods of growth followed by recessions, followed in turn by expansions—remains an economic (and political) fact of life.

Seasonal variations are shorter demand cycles that depend on the time of year. Seasonal factors affect tourism and air travel, tax preparation services, clothing, and other products and services.

Finally, one should not ignore the role of **random fluctuations**. In any short period of time, an economic variable may show irregular movements due to essentially random (or unpredictable) factors. For

instance, a car dealership may see 50 more customers walk into its showroom one week than the previous week and, therefore, may sell eight more automobiles. Management is grateful for the extra sales even though it can identify absolutely no difference in economic circumstances between the two weeks. Random fluctuations and unexpected occurrences are inherent in almost all time series. No model, no matter how sophisticated, can perfectly explain the data.

Fitting a Simple Trend

4.4 plots the level of annual sales for a product over a dozen years. The time series displays a smooth upward trend. One of the simplest methods of time-series forecasting is fitting a trend to past data and then extrapolating the trend into the future to make a forecast. Let's first estimate a linear trend, that is, a straight line through the past data.

Barometric Models

Barometric models search for patterns among different variables over time. Consider a firm that produces oil drilling equipment. Management naturally would like to forecast demand for its product. It turns out that the seismic crew count, an index of the number of teams surveying possible drilling sites, gives a good indication as to changes in future demand for drilling equipment. For this reason, we call the seismic crew a **leading indicator** of the demand for drilling equipment.

Economists have identified many well-known leading indicators. The number of building permits lead the number of housing starts. Stock market indices (such as the Dow Jones Industrial Average) indicate future increases and decreases in economic activity (expansions or recessions). Such indicators, however, are not without certain problems.

1. Leading indicators are not always accurate. According to one humorous economic saying, declines in the stock market have predicted 14 of the last 8 recessions.
2. The amount of time between the change in the leading indicator and the change in the forecasted series varies. Leading indicators may say a change is coming, but they often cannot say exactly when.
3. The change in the leading indicator rarely gives much information about the size of the change in the forecasted series.

Frequently, leading indicators are averaged to form a composite leading indicator. This helps eliminate some of the randomness and makes the indicator 162 Unit 4 Estimating and Forecasting Demand more accurate. The U.S. Bureau of Economic Analysis has developed (and publishes) the **Index of Leading Indicators**. This index signals future changes in the course of the economy. The revised index is a weighted average of 11 economic series:

1. Weekly hours of manufacturing workers
2. Manufacturers' new orders
3. Changes in manufacturers' unfilled orders
4. Plant and equipment orders
5. The number of housing building permits
6. Changes in sensitive materials prices
7. Percentage of companies receiving slower deliveries
8. The money supply
9. The index of consumer confidence

10. The index of 500 companies' common-stock prices
11. Average weekly claims for unemployment insurance

Positive changes in the first 10 indicators (and a decline in the last) indicate future economic growth, whereas persistent declines in the index presage a weak economy and possible recession. On average, the composite index tends to turn down nine months before the onset of recession. The index increases about four to five months before the economy bottoms out and begins to grow.

REVIEW QUESTION

1. Discuss and compare the advantages and disadvantages of survey methods and test marketing.

1. Describe the process of collecting data. What are the survey pitfalls ?
2. Discuss about controlled market studies. What are uncontrolled market data?
3. Describe regression analysis and r-squared .
4. What is equation specification and omitted variables
5. Describe multi-collinearity and simultaneity and identification
6. Discuss about forecasting? Explain about time-series models.

FURTHER READINGS

1. Managerial Economics - Arun Kumar, Rachana Sharma
2. Managerial Economics - Thomas J. Webster
3. Managerial Economics - Petersen / Jain
4. Managerial Economics - Yogesh Maheshwari
5. Managerial Economics - E. Narayanan Nadar, S. Vijayan

UNIT-5 PRODUCTION

CONTENTS

INTRODUCTION

We open this Unit by examining the production function, a quantitative summary of the firm's production possibilities. Next, we look closely at production in the short run and examine the impact on output of changing a single input. Then we consider production in the long run, when the firm has the flexibility to vary the amounts of all inputs. Next, we turn to the various types of production functions and discuss the means by which they are estimated. Finally, we consider a number of constrained production decisions involving the allocation of inputs (in fixed supply) to multiple plants or products, or both.

BASIC PRODUCTION CONCEPTS

Production transforms inputs into outputs. For instance, producing automobiles requires a variety of inputs (also called factors of production): raw materials (steel, plastic, rubber, and so on), factories, machines, land, and many different categories of workers. For analysis, it is convenient to refer to two main categories of inputs—labor and materials on the one hand and long-term capital on the other—with each category broadly defined. Labor and materials

includes production workers, marketers, and managers at all levels as well as raw materials and intermediate goods, including parts, water, and electricity. Capital includes buildings, equipment, and inventories.

The firm's **production function** indicates the maximum level of output the firm can produce for any combination of inputs. We will start by considering a production function with two inputs, labor and capital. This states that the firm's quantity of output depends on the respective quantities of labor (L) and capital (K). For instance, a major domestic automobile manufacturer might plan to produce 3 million passenger cars per year, using materials (of all kinds) that cost $24 billion, a total nationwide labor force of 80,000 workers, and a total capital stock valued at $100 billion. Note that the firm's production function specifies the maximum output for a given combination of inputs. It assumes that managers use inputs efficiently. Obviously, production technologies improve over time, and efficient firms vigorously pursue these improvements.

$$Q = F(L, K).$$

PRODUCTION WITH ONE VARIABLE INPUT

Short-Run and Long-Run Production

Our analysis of production and cost makes an important distinction between the short run and the long run.

In the **short run** one or more of the firm's inputs is fixed; that is, they cannot be varied. In the **long run** the firm can vary all of its inputs. There is no universal rule for distinguishing between the short and long run; rather, the dividing line must be drawn on a case-by-case basis. For a petrochemical refinery, the short run might be any period less than five years since it takes roughly this long to build a new refinery. For a fast-food chain, six months (the time it takes to obtain zoning approvals and construct new restaurants) may be the dividing line between the short and long run.

Inputs that cannot be changed in the short run are called **fixed inputs**. A firm's production facility is a typical example. In the long run, the firm could vary the size and scale of its plant, whereas in the short run the size of this plant would be fixed at its existing capacity. If a firm operates under restrictive, longterm labor contracts, its ability to vary its labor force may be limited over the contract duration, perhaps up to three years. In this case, labor could be a fixed input in the short run.

MARGINAL PRODUCT Let's consider the production decisions of the auto parts firm. Currently it is operating with a 10,000-square-foot plant. In the short run, this capital input is fixed. However, labor is a **variable input**; that is, the firm can freely vary its number of workers.

THE LAW OF DIMINISHING MARGINAL RETURNS The declining marginal product of an input (like labor) represents one of the best-known and most important empirical "laws" of production:

The Law of Diminishing Marginal Returns. As units of one input are added (with all other inputs held constant), resulting additions to output will eventually begin to decrease; that is, marginal product will decline.

In the preceding example, diminishing returns to labor occur beyond 40 workers. At this point the most productive jobs already are filled, specialization is being fully exploited, and the plant and equipment are being used efficiently.

Optimal Use of an Input

The law of diminishing returns means that the firm faces a basic trade-off in determining its level of production. By using more of a variable input, the firm obtains a direct benefit—increased output—in return for incurring an additional input cost. What level of the input maximizes profits? As before, we look at the firm's marginal profit, but this time we look at marginal profit per unit of input. We increase the input until the marginal profit per unit of input is zero.

In analyzing this input decision, a definition is helpful. Marginal revenue product is the formal name for the marginal revenue associated with increased use of an input. An input's **marginal revenue product** is the extra revenue that results from a unit increase in the input. To illustrate,

suppose the auto parts supplier is considering increasing labor from 20 to 30 workers.

PRODUCTION IN THE LONG RUN

In the long run, a firm has the freedom to vary all of its inputs. Two aspects of this flexibility are important. First, a firm must choose the proportion of inputs to use. For instance, a law firm may vary the proportion of its inputs to economize on the size of its clerical staff by investing in computers and software specifically designed for the legal profession. In effect, it is substituting capital for labor. Steeply rising fuel prices have caused many of the major airlines to modify their fleets, shifting from larger aircraft to smaller, fuel-efficient aircraft.

Second, a firm must determine the scale of its operations. Would building and operating a new facility twice the size of the firm's existing plants achieve a doubling (or more than doubling) of output? Are there limits to the size of the firm beyond which efficiency drastically declines? These are all important questions that can be addressed using the concept of returns to scale.

Returns to Scale

The scale of a firm's operations denotes the levels of all the firm's inputs. A change in scale refers to a given percentage change in all inputs. At a 15 percent scale increase, the firm would use 15 percent more of each of its inputs. A key question for the manager is how the change in scale affects the firm's output. **Returns to scale** measure the percentage change in output resulting from a given percentage change in inputs. There are three important cases.

Constant returns to scale occur if a given percentage change in all inputs results in an equal percentage change in output. For instance, if all inputs are doubled, output also doubles; a 10 percent increase in inputs results in a 10 percent increase in output; and so on. A common example of constant returns to scale occurs when a firm can easily replicate its production process. For instance, a manufacturer of electrical components might find that it can double its output by replicating its current plant and labor force, that is, by building an identical plant beside the old one.

Increasing returns to scale occur if a given percentage increase in all inputs results in a greater percentage change in output. For example, a 10 percent increase in all inputs causes a 20 percent increase in output. How can the firm do better than constant returns to scale? By increasing its scale, the firm may be able to use new production methods that were infeasible at the smaller scale. For instance, the firm may utilize sophisticated, highly efficient, large-scale factories. It also may find it advantageous to exploit specialization of labor at the larger scale. As an example, there is considerable evidence of increasing returns to scale in automobile manufacturing: An assembly plant with a capacity of 200,000 cars per year uses significantly less than twice the input quantities of a plant having a 100,000-car capacity. Frequently, returns to scale result from fundamental engineering relationships. Consider the economics of an oil pipeline from well sites in Alaska to refineries in the contiguous United States. Doubling the circumference of the pipe increases the pipe's cross sectional area fourfold—allowing a like

increase in the flow capacity of the pipeline. To sum up, as long as there are increasing returns, it is better to use larger production facilities to supply output instead of many smaller facilities.

Decreasing returns to scale occur if a given percentage increase in all inputs results in a smaller percentage increase in output. The most common explanations for decreasing returns involve organizational factors in very large firms. As the scale of the firm increases, so do the difficulties in coordinating and monitoring the many management functions. As a result, proportional increases in output require more than proportional increases in inputs.

Output elasticity is the percentage change in output resulting from a 1 percent increase in all inputs. For constant returns to scale, the output elasticity is 1; for increasing returns, it is greater than 1; and for decreasing returns, it is less than 1. For instance, an output elasticity of 1.5 means that a 1 percent scale increase generates a 1.5 percent output increase, a 10 percent scale increase generates a 15 percent output increase, and so on.

Least-Cost Production

In the long run, the firm can vary all of its inputs. Because inputs are costly, this flexibility raises the question: How can the firm determine the mix of inputs that will minimize the cost of producing a given level of output? To answer this question, let's return to the case of two inputs, labor and capital. Here the firm's production function is of the form where L is the number of labor hours per month and K is the amount of capital used per month. There are possibly many different ways to produce a given level of output (call this Q_0), utilizing more capital and less labor or vice versa. The optimal mix of labor and capital in producing output Q_0 depends on the costs and marginal products of the inputs. Let's denote the firm's labor

cost per hour by P_L and its cost per unit of capital by P_K. Then the firm's total cost of using L and K units of inputs is

MEASURING PRODUCTION FUNCTIONS

In this section, we briefly discuss ways in which managers can estimate and measure production functions based on engineering or economic data. Let us begin by considering four common specifications.

Linear Production

As the term suggests, a **linear production function** takes the form where a, b, and c are coefficients that must be estimated from the data. An immediate implication of linearity is that each input's marginal product is constant: MP_L _ a and MP_K _ b. Constant marginal productivity may approximate production over a limited range of input usage, but at sufficiently high levels of inputs, it is at odds with the law of diminishing marginal productivity. In this sense, the linear form is too simple and should be viewed as a somewhat extreme case. Because of the constant marginal products, the inputs are perfect substitutes for one another. Suppose, for example, that the production function is Q _ 20L _ 40K. In this case, one can always substitute two units of labor for one of capital to maintain the same level of production, and vice versa. Given fixed input prices, production will be "all or nothing" in the long run. If

the unit cost of capital is less than twice the wage per unit of labor, the firm's least-cost means of production is to use only capital. In contrast, if labor is the less expensive option, production should use labor exclusively. In general, as long as MPK/PK _ MPL/PL, the firm should use capital exclusively (and vice versa if the inequality is reversed).

Production with Fixed Proportions

Production with fixed proportions is the opposite extreme from linear production; fixed-proportions production allows no input substitution. Output can only be produced with a fixed proportion of inputs. Simple examples include a taxi and its driver or a construction crane and its operator. In both cases, the required mix of labor to capital is one to one. An excess of either input—a machine without an operator or vice versa—does no good. Expansion of production requires balanced increases in the necessary inputs. Like linear production, fixed-proportions production should be thought of as an extreme case. Rarely is there no opportunity for input substitution. (For example, it is true that a crane needs an operator but, at a more general level, extra construction workers can substitute for construction equipment.) However, fixed-proportions production has an important implication. In the face of an increase in an input's price, the firm cannot economize on its use, that is, substitute away from it. Thus, a petrochemical firm that uses fixed proportions of different chemicals to produce its specialty products is at the mercy of market forces that drive up the prices of some of these inputs.

Polynomial Functions

In the polynomial form, variables in the production function are raised to positive integer powers. As a simple example, consider the quadratic form where a and b are positive coefficients. It is easy to check that each input shows diminishing returns. (For example, MPL $Q = aLK - bL^2K^2$, , which declines as L increases.) The quadratic form also displays decreasing returns to scale. A more flexible representation is the cubic form, where all coefficients are positive. We can show that this function displays increasing returns for low-output levels and then decreasing returns for highoutput levels. The marginal product of an input (say, labor) takes the form We see that marginal product is a quadratic function in the amount of labor; that is, it is a parabola that rises, peaks, and then falls. Thus, this functional form includes an initial region of increasing marginal productivity followed by diminishing returns.

Estimating Production Functions

Data for estimating production functions come in a number of forms. Engineering data can provide direct answers to a number of production questions: On average, how much output can be produced by a certain type of machine under different operating conditions? How many bushels of a particular crop can be grown and harvested on land (of known quality) using specified amounts of labor, capital, and materials (such as fertilizer)? Such information usually is based on experience with respect to similar (or not so similar) production processes. Consequently, the estimated production function is only as accurate as the past

production experience on which it is based. The development of new weapons systems is a case in point. Although production and cost estimates are based on the best available engineering estimates (and possibly on tests of prototypes), they nonetheless are highly uncertain.9

A second source of production information is production data. For example, in a production time-series analysis, the firm's managers compile a production history, month by month or year by year, recording the amounts of inputs (capital, labor, land, materials, and so on) used in production and the resulting level of output. Alternatively, the economic data may come in the form of a cross section. In this case, information is gathered for different plants and firms in a given industry during a single period of time. For instance, by observing production in the auto industry, one can address a number of important questions: For plants of fixed size (possibly employing different degrees of automation), what is the effect on output of expanding the labor force (for instance, adding extra shifts)? Does the industry exhibit economies of scale and, if so, over what range of outputs? (That is, will a 40 percent increase in plant scale deliver more than a 40 percent increase in output?)

Production data—though subject to measurement errors—are very useful to managers. Based on these data, the manager (often with the help of an operations research specialist) can estimate the mathematical relationship between levels of inputs and quantity of output. The principal statistical method for carrying out this task is regression analysis (the most important elements of which were discussed in Unit 4). The end product of this analysis is a tangible representation of the firm's production function.

OTHER PRODUCTION DECISIONS

Within the limits of its production technology, the firm's managers face a number of important decisions. We have already discussed finding the optimal use of single input in the short run and choosing the best mix of inputs in the long run. We now consider two other decisions: (1) the allocation of a single input among multiple production facilities and (2) the use of an input across multiple products.

Multiple Plants

Consider an oil company that buys crude oil and transforms it into gasoline at two of its refineries. Currently it has 10 thousand barrels of oil under long-term contract and must decide how to allocate it between its two refineries. The company's goal is to allocate supplies to maximize total output from the refineries. Let M_A and M_B represent the crude input at each refinery and Q_A and Q_B the gasoline outputs. The firm's problem is:

$$\text{Maximize } Q = Q_A + Q_B, \text{ subject to } M_A + M_B = 10 \text{ thousand.}$$

The key to maximizing total output is to compare marginal products at the two refineries. Barrels of crude first should be allocated to the refinery at which the marginal product is greater. Let's say this is refinery A. As additional barrels are allocated to this refinery, its marginal product diminishes, and it becomes worthwhile to allocate oil to refinery B as well.

In the final allocation of all 10 thousand barrels, output is maximized if and only if the marginal products of both refineries are equal, that is, when $MP_A = MP_B$. Maximize Q _ QA _ QB, subject to

Multiple Products

Firms often face the problem of allocating an input in fixed supply among different products. The input may be a raw material—for instance, DRAM computer chips allocated to the various models of personal computers manufactured by the firm—or it may be capital. Frequently, the input in shortest supply is managerial labor itself. Which products of the firm are in greatest need of managerial attention? Which top-level managers are best suited to improve performance in a given product line?

REVIEW QUESTION

Questions and Problems

1. Does optimal use of an input (such as labor) mean maximizing average output (per unit of input)? Explain.
2. "One-tenth of the participants produce over one-third of the output. Increasing the number of participants merely reduces the average output." If this statement were true, would it be consistent with the law of diminishing returns?
3. Explain the difference between diminishing returns and decreasing returns to scale.
4. Describe basic production concepts.
5. What is the production with one variable input?
6. Explain the law of diminishing marginal returns.
7. Discuss about production in the long run.
8. Describe the process measuring production functions.

FURTHER READINGS

1. Managerial Economics - Arun Kumar, Rachana Sharma
2. Managerial Economics - Thomas J. Webster
3. Managerial Economics - Petersen / Jain
4. Managerial Economics - Yogesh Maheshwari
5. Managerial Economics - E. Narayanan Nadar, S. Vijayan

UNIT-6 COST ANALYSIS

CONTENTS

INTRODUCTION

In this Unit, we build on Unit 5's analysis of production to provide an overview of these crucial cost concepts. In the first section, we discuss the basic principles of relevant costs—considering the concepts of opportunity costs and fixed costs in turn. Next, we examine the relationship between cost and output in the short run and the long run. Then we turn to economies of scale and economies of scope. Finally, we consider the importance of cost analysis for a number of key managerial decisions.

RELEVANT COSTS

A continuing theme of previous Units is that optimal decision making depends crucially on a comparison of relevant alternatives. Roughly speaking, the manager must consider the relevant pros and cons of one alternative versus another. The precise decision-making principle is as follows: In deciding among different courses of action, the manager need only consider the differential revenues and costs of the alternatives.

Thus, the only relevant costs are those that differ across alternative courses of action. In many managerial decisions, the pertinent cost differences are readily apparent. In others, issues of relevant cost are more subtle. The notions of opportunity costs and fixed costs are crucial for managerial decisions. We will consider each topic in turn.

Opportunity Costs and Economic Profits

The concept of opportunity cost focuses explicitly on a comparison of relative pros and cons. The **opportunity cost** associated with choosing a particular decision is measured by the benefits forgone in the next-best alternative. Typical examples of decisions involving opportunity cost include the following:

• What is the opportunity cost of pursuing an MBA degree?

• What is the opportunity cost of using excess factory capacity to supply specialty orders?

• What is the opportunity cost that should be imputed to city-owned land that is to be the site of a public parking garage downtown?

As the definition suggests, an estimate of the opportunity cost in each case depends on identifying the next-best alternative to the current

decision. Consider the first example. Suppose the MBA aspirant has been working in business for five years. By pursuing an MBA degree full time, what is he giving up? Presumably, it is the income he could have earned from the present job. (This opportunity cost is larger or smaller depending on how remunerative the job is and on the chances for immediate advancement.) Therefore, the total cost of taking an MBA degree is the explicit, out-of-pocket tuition cost plus the implicit (but equally real) opportunity cost.1

ECONOMIC PROFIT

At a general level, the notion of profit would appear unambiguous: Profit is the difference between revenues and costs. On closer examination, however, one must be careful to distinguish between two definitions of profit. **Accounting profit** is the difference between revenues obtained and expenses incurred. The profit s reported by firms almost always are based on accounting profits; it is the job of accountants to keep a careful watch on revenues and explicit expenses. This information is useful for both internal and external purposes: for managers, shareholders, and the government (particularly for tax purposes). With respect to managerial decision making, however, the accounting measure does not present the complete story concerning profitability. In this case, the notion of economic profit is essential. **Economic profit** is the difference between revenues and all economic costs (explicit and implicit), including opportunity costs. In particular, economic profit involves costs associated with capital and with managerial labor. Here is a simple illustration.

Fixed and Sunk Costs

Costs that are **fixed**—that is, do not vary—with respect to different courses of action under consideration are irrelevant and need not be considered by the manager. The reason is simple enough: If the manager computes each alternative's profit (or benefit), the same fixed cost is subtracted in each case. Therefore, the fixed cost itself plays no role in determining the relative merits of the actions. Consider once again the recent graduate who is deciding whether to begin work immediately or to take an MBA degree. In his deliberations, he is concerned about the cost of purchasing his first car. Is this relevant?

The answer is no, assuming he will need (and will purchase) a car whether he takes a job or pursues the degree. Consider a typical business example. A production manager must decide whether to retain his current production method or switch to a new method. The new method requires an equipment modification (at some expense) but saves on the use of labor. Which production method is more profitable? The hard (and tedious) way to answer this question is to compute the bottom-line profit for each method. The easier and far more insightful approach is to ignore all fixed costs. The original equipment cost, costs of raw materials, selling expenses, and so on are all fixed (i.e., do not vary) with respect to the choice of production method. The only differential costs concern the equipment modification and the reduction in labor. Clearly, the new method should be chosen if and only if its labor savings exceed the extra equipment cost. Notice that the issue of relevant costs would be very different if management were tackling the larger decision of whether to

continue production (by either method) or shut down. With respect to a shut-down decision, many (if not all) of the previous fixed costs become variable. Here the firm's optimal decision depends on the magnitudes of costs saved versus revenues sacrificed from discontinuing production.

Ignoring fixed costs is important not only because it saves considerable computation but also because it forces managers to focus on the differential costs that are relevant. Be warned that ignoring fixed costs is easier in principle than in practice. The case of sunk costs is particularly important.

A **sunk cost** is an expense that already has been incurred and cannot be recovered. For instance, in the earlier factory example, plant space originally may have been built at a high price. But this historic cost is sunk and is irrelevant to the firm's current decision. As we observed earlier, in the case of excess, unused factory capacity, the relevant opportunity cost is near zero.

More generally, sunk costs cast their shadows in sequential investment decisions. Consider a firm that has spent $20 million in research and development on a new product. The R&D effort to date has been a success, but an additional $10 million is needed to complete a prototype product that (because of delays) may not be first to market. Should the firm make the additional investment in the product? The correct answer depends on whether the product's expected future revenue exceeds the total additional costs of developing and producing the product. (Of course, the firm's task is to forecast accurately these future revenues and costs.) The $20 million sum spent to date is sunk and, therefore, irrelevant for the firm's decision. If the product's future prospects are unfavorable, the firm should cease R&D.

Perhaps the last word on sunk cost is provided by the story of the seventeenth century warship Vassa. When newly launched in Stockholm before a huge crowd that included Swedish royalty, the ship floated momentarily, overturned, and ignominiously (and literally) became a sunk cost.

Profit Maximization with Limited Capacity: Ordering a Best Seller

The notion of opportunity cost is essential for optimal decisions when a firm's multiple activities compete for its limited capacity. Consider the manager of a bookstore who must decide how many copies of a new best seller to order. Based on past experience, the manager believes she can accurately predict potential sales. Suppose the best seller's estimated price equation is P _24_ Q, where P is the price in dollars and Q is quantity in hundreds of copies sold per month. The bookstore buys directly from the publisher, which charges $12 per copy. Let's consider the following three questions:

1. How many copies should the manager order, and what price should she charge? (There is plenty of unused shelf space to stock the best seller.)

2. Now suppose shelf space is severely limited and stocking the best seller will take shelf space away from other books. The manager estimates that there is a $4 profit on the sale of a book stocked. (The best

seller will take up the same shelf space as the typical book.) Now what are the optimal price and order quantity?

3. After receiving the order in Question 2, the manager is disappointed to find that sales of the best seller are considerably lower than predicted. Actual demand is P _ 18 _ 2Q. The manager is now considering returning some or all of the copies to the publisher, who is obligated to refund $6 for each copy returned. How many copies should be returned (if any), and how many should be sold and at what price? As always, we can apply marginal analysis to determine the manager's optimal course of action, provided we use the "right" measure of costs. In out, other government programs, once begun, seem to have lives of their own.

Question 1, the only marginal cost associated with the best seller is the explicit $12 cost paid to the publisher. The manager maximizes profit by setting MR equal to MC. Since MR _ 24 _ 2Q, we have 24 _ 2Q _ 12. The result is Q _ 6 hundred books and P _ $18. This outcome is listed in Table 6.1a. By comparison, what are the optimal order quantity and price when shelf space is limited, as in Question 2? The key point is that ordering an extra best seller will involve not only an out-of-pocket cost ($12) but also an opportunity cost ($4). The opportunity cost is the $4 profit the shelf space would earn on an already stocked book—profit that would be forgone. In short, the total cost of ordering the book is 12 _ 4 _ $16. Setting MR equal to $16, we find that Q _ 4 hundred and P _ $20. Given limited shelf space, the manager orders fewer best sellers than in Question 1. Table 6.1b compares the profitability of ordering 400 versus 600 books. The cost column lists the store's payment to the publisher ($12 per best seller). Forgone profit is measured at $4 per book.

We confirm that ordering 400 books is the more profitable option, taking into account the forgone profit on sales of other books. Indeed, the logic of marginal analysis confirms that this order quantity is optimal, that is, better than any other order size.

Finally, Question 3 asks how the manager should plan sales and pricing of the 400 best sellers already received if demand falls to P _ 18 _ 2Q. The key here is to recognize that the original $12 purchase price is irrelevant; it is a sunk cost. However, opportunity costs are relevant. The opportunity cost of keeping the best seller for sale has two elements: the $4 profit that another book would earn (as in Question 2) plus the $6 refund that would come if the copy were returned. Therefore, the total opportunity cost is 6 _ 4 _ $10.

THE COST OF PRODUCTION

As we noted in Unit 5, production and cost are very closely related. In a sense, cost information is a distillation of production information: It combines the information in the production function with information about input prices. The end result can be summarized in the following important concept: The **cost function** indicates the firm's total cost of producing any given level of output. The concept of a cost function was first introduced in Unit 2. In this section, we take a much closer look at the factors that determine costs. A key

point to remember is that the concept of the cost function presupposes that the firm's managers have determined the least-cost method of producing any given level of output. (Clearly, inefficient or incompetent

managers could contrive to produce a given level of output at some—possibly inflated—cost, but this would hardly be profit maximizing. Nor would the resulting cost schedule foster optimal managerial decision making.) In short, the cost function should always be thought of as a least-cost function. It usually is denoted as C _ C(Q) and can be described by means of a table, a graph, or an equation. As in our study of production, our analysis of cost distinguishes between the short run and the long run. Recall that the short run is a period of time so limited that the firm is unable to vary the use of some of its inputs. In the long run, all inputs—labor, equipment, factories—can be varied freely. Our investigation of cost begins with the short run.

Short-Run Costs

In the basic model of Unit 5, we focused on two inputs, capital and labor. In the short run, capital is a fixed input (i.e., cannot be varied) and labor is the sole variable input. Production of additional output is achieved by using additional hours of labor in combination with a fixed stock of capital equipment in the firm's current plant. Of course, the firm's cost is found by totaling its expenditures on labor, capital, materials, and any other inputs and including any relevant opportunity costs, as discussed in the previous section. For concreteness, consider a firm that provides a service—say, electronic repair. 6.1 provides a summary of the repair firm's costs as they vary for different quantities of output (number of repair jobs completed).

The total cost of achieving any given level of output can be divided into two parts: fixed and variable costs. As the term suggests, **fixed costs** result from the firm's expenditures on fixed inputs. These costs are incurred regardless of the firm's level of output. Most overhead expenses fall into this category. Such costs might include the firm's lease payments for its factory, the cost of equipment, some portion of energy costs, and various kinds of administrative costs (payment for support staff, taxes, and so on). According to the table in 6.1, the repair firm's total fixed costs come to $270,000 per year. These costs are incurred regardless of the actual level of output (i.e., even if no output were produced).

Variable costs represent the firm's expenditures on variable inputs. With respect to the short-run operations of the repair firm, labor is the sole variable input. Thus, in this example, variable costs represent the additional wages paid by the firm for extra hours of labor. To achieve additional output (i.e., to increase the volume of repair jobs completed), the firm must incur additional variable costs. Naturally, we observe that total variable costs rise with increases in the quantity of output. In fact, a careful look at 6.1 shows that variable costs rise increasingly rapidly as the quantity of output is pushed higher and higher. Note that the firm's total cost exhibits exactly the same behavior. (With fixed costs "locked in" at $270,000, total cost increases are due solely to changes in variable cost.) The graph in 6.1 shows that the total cost curve becomes increasingly steep at higher output levels.

Average total cost (or simply **average cost**) is total cost divided by the total quantity of output. 6.2 shows average costs for the repair company over different levels of output. (Check that the average cost values are

computed as the ratio of total cost in column 2 of the table and total output in column 1.) The graph displays the behavior of average cost. Both the table and graph show that short-run average cost is U-shaped. Increases in output first cause average cost (per unit) to decline. At 30,000 units of output, average cost achieves a minimum (at the bottom of the U). As output continues to increase, average unit costs steadily rise. (We will discuss the factors underlying this average cost behavior shortly.) Finally, **average variable cost** is variable cost divided by total output. Because it excludes fixed costs, average variable cost is always smaller than average total cost.

Marginal cost is the addition to total cost that results from increasing output by one unit. We already are acquainted with the concept of marginal cost from the analyses of the firm's output and pricing decisions in Units 2 and 3. Now we take a closer look at the determinants of marginal cost. The last column of the table in 6.2 lists the repair company's marginal costs for output increments of 5,000 units. For instance, consider an output increase from 25,000 to 30,000 units. According to 6.2, the result is a total cost increase of 1,440,000 _ 1,207,500 _ $232,500. Consequently, the marginal cost (on a perunit basis) is 232,500/5,000 _ $46.50/unit. The other entries in the last column are computed in an analogous fashion. From either the graph or the table, we observe that the firm's marginal cost rises steadily with increases in output. Expanding output starting from a level of 40,000 units per month is much more expensive than starting from 20,000 units. What factors underlie the firm's increasing short-run marginal cost (SMC)? The explanation is simple. With labor the only variable input, SMC can be expressed as where P_L denotes the price of hiring additional labor (i.e., wage per hour) and MP_L denotes the marginal product of labor.[5] To illustrate, suppose the prevailing wage is $20 per hour and labor's marginal product is .5 unit per hour (one-half of a typical repair job is completed in one hour). Then the firm's marginal (labor) cost is 20/.5 _ $40 per additional completed job. According to Equation 6.1, the firm's marginal cost will increase if there is an increase in the price of labor or a decrease in labor's marginal product. Moreover, as the firm uses additional labor to produce additional output, the law of diminishing returns applies. With other inputs fixed, adding increased amounts of a variable input (in this case, labor) generates smaller amounts of additional output; that is, after a point, labor's marginal product declines. As a result, marginal cost rises with the level of output. (Clearly, material costs are also variable and, therefore, are included in SMC. However, because these costs typically vary in proportion to output, they do not affect the shape of SMC.) Now we can explain the behavior of short-run average cost (SAC). When output is very low (say 5,000 units), total cost consists mainly of fixed cost (since variable costs are low). SAC is high because total cost is divided by a small number of units. As output increases, total costs (which are mostly fixed) are "spread over" a larger number of units, so SAC declines. In the graph in 6.2, notice that SAC lies well above SMC for low levels of output. As long as extra units can be added at a marginal cost that is lower than the average cost of the current output level, increasing output must reduce overall average cost.

But what happens to average cost as marginal cost continues to rise? Eventually there comes a point at which SMC becomes greater than SAC. As soon as extra units become more expensive than current units (on average), the overall average begins to increase. This explains the upward arc of the U-shaped SAC curve. This argument also confirms an interesting result: The firm's marginal cost curve intersects its average cost curve at the minimum point of SAC.

Long-Run Costs

In the long run, the firm can freely vary all of its inputs. In other words, there are no fixed inputs or fixed costs; all costs are variable. Thus, there is no difference between total costs and variable costs. We begin our discussion by stressing two basic points. First, the ability to vary all inputs allows the firm to produce at lower cost in the long run than in the short run (when some inputs are fixed). In short, flexibility is valuable. As we saw in Unit 5, the firm still faces the task of finding the least-cost combination of inputs.Second, the shape of the long-run cost curve depends on returns to scale.To see this, suppose the firm's production function exhibits constant returns to scale. **Constant returns to scale** means that increasing all inputs by a given percentage (say, 20 percent) increases output by the same percentage.

Assuming input prices are unchanged, the firm's total expenditure on inputs also will increase by 20 percent. Thus, the output increase is accompanied by an equal percentage increase in costs, with the result that average cost is unchanged.

SHORT-RUN VERSUS LONG-RUN COST

Consider a firm that produces output using two inputs, labor and capital. Management's immediate task is to plan for future production. It has not leased plant and equipment yet, nor has it hired labor. Thus, it is free to choose any amounts of these inputs it wishes. Management knows that production exhibits constant returns to scale. Consequently,

the firm's long-run average cost (LAC) is constant as shown by the horizontal line in 6.3. Furthermore, we can show that the firm should plan to use the same optimal ratio of labor to capital in production, regardless of the level of output. If the firm plans to double its level of output, it should also double the use of e ch input, leaving the proportions unchanged.

RETURNS TO SCALE AND SCOPE

Returns to Scale

Returns to scale are important because they directly determine the shape of long-run average cost. They also are crucial for answering such questions as Are large firms more efficient producers than small firms? Would a 50 percent increase in size reduce average cost per unit? Although the exact nature of returns to scale varies widely across industries, a representative description is useful. 6.4 depicts a long-run average cost curve that is U-shaped. This reflects increasing returns to scale (and falling LAC) for low output levels and decreasing returns (increasing LAC) for high levels. In the , the minimum level of long-run average cost is achieved at output level Q_{min}. As in 6.3, SAC curves for three plants are shown. Thus, output Q_{min} is produced using the medium-

sized plant. If the costs of all possible plants were depicted, the lower "envelope" of the many SAC curves would trace out the's LAC curve. To sum up, if the firm is free to use any size plant, its average production cost is exactly LAC.

As noted in Unit 5, a number of factors influence returns to scale and, therefore, the shape of long-run average cost. First, constant average cost (due to constant returns to scale) occurs when a firm's production process can be replicated easily. For instance, the electronics repair firm may find it can double its rate of finished repair jobs simply by replicating its current plant and labor force—that is, by building an identical repair facility beside the existing one and proportionally increasing its labor force. By duplication, the firm could supply twice the level of service at an unchanged average cost per job. Second, declining average cost stems from a number of factors, including capital-intensive mass production techniques, automation, labor specialization, advertising, and distribution. By increasing scale, the firm may be able to use new production methods that were infeasible at smaller outputs. It also may find it advantageous to exploit specialization of labor at the larger scale. The result of either kind of production innovation is a reduction in long-run average cost.

Fundamental engineering relationships may have the same effect. For instance, in 2011, Royal Caribbean International boasted the world's largest cruise liner, costing $1.1 billion, with capacity for 6,400 passengers and 2,300 crew. The largest cruise ships take full advantage of scale economies. At twice the tonnage, a super-cruise liner can carry significantly more than twice the number of passengers while requiring only a relatively modest increase in crew. Accordingly, the cost per passenger declines markedly.

Declining average cost also may be due to the presence of a variety of fixed expenses. Frequently, significant portions of a firm's advertising, promotional, and distributional expenses are fixed or (at least) vary little with the firm's level of output. (For instance, a 30-second television advertisement represents the same fixed cost to a large fast-food chain and a small chain alike. But this expense constitutes a much lower average cost per burger for the large chain.) Similarly, the costs to firms of many government regulations are (in the main) fixed. Accordingly, they represent a smaller average cost for the large firm. The U.S. automobile industry, perhaps the most highly regulated sector in the world, is a case in point.

Finally, increasing average cost is explained by the problems of organization, information, and control in very large firms. As the firm's scale increases, so do the difficulties of coordinating and monitoring its many management functions. The result is inefficiency, increased costs, and organizational overload.6 A great many studies have investigated the shape of average cost curves for different industries in both the short and long runs. Almost all of these studies use regression techniques to generate equations that explain total cost as a function of output and other relevant explanatory variables (such as wages and other input prices). The data for this analysis can come from either a time series (the same firm over a number of months or years) or a cross section (a cost

comparison of different firms within a single time period). Despite difficulties in estimating costs from accounting data and controlling for changing inputs (especially capital), technology, and product characteristics, these studies have produced valuable information about costs.

E-Commerce and Cost Economies

As noted in Unit 3, the Internet and the emergence of e-commerce have significant impacts on the structure of firm costs.8 A wide-ranging research study by Washington's Brookings Institution estimated that across the whole of the U.S. economy, the adoption of information technology and e-commerce methods was producing total annual cost savings of a magnitude equivalent to about 1 percent of annual gross domestic product. Increased efficiency stemmed from reengineering the firm's supply chain and from reducing transactions costs of all kinds. The greatest potential savings emerged in information intensive industries such as health care, financial services, education, and public-sector operations.

Recall that the hallmark of information economics is the presence of high fixed costs accompanied by low or negligible marginal costs. As a result, average costs decline sharply with output. The fixed costs of business capital investments are increasingly found in computers, computing systems such as servers, software, and telecommunications (together constituting over 10 percent of capital expenditure), rather than in the traditional "bricks and mortar" of factories, assembly lines, and equipment. To date, a number of firms—Microsoft, Google, Cisco Systems, Oracle, eBay, Facebook, and YouTube, to name a few—have taken advantage of information economies to claim increasing shares of their respective markets, thus, benefiting from sharply declining average unit costs.

E-commerce also benefits from significant economies of scale in customer acquisition and service. In many e-commerce markets there has been a land-rush-like frenzy to acquire customers (often by offering a variety of free services). These customers come at a high initial fixed cost but have a very low marginal cost of servicing them. In addition, demand-side externalities mean that customers receive greater value as the population of other customers increase. This is true in online sites ranging from job-search to business-to-business commerce to online classified ads. For instance, such economies of scale provide eBay and Google with dominant positions in online auctions and search, respectively. In turn, economies of scale in distribution means that at large enough scale, taking orders online, holding inventories in centralized facilities, and shipping direct to customers is cheaper than selling the same item at a retail outlet. The online sales clout of Amazon is an obvious case in point.

Economies of Scope

Most firms produce a variety of goods. Computer firms, such as IBM and Toshiba, produce a wide range of computers from mainframes to personal computers. Consumer products firms, such as Procter & Gamble and General Foods, offer myriad personal, grocery, and household items. Entertainment firms, such as Walt Disney Corporation,

produce movies, television programs, toys, theme park entertainment, and vacation services. In many cases, the justification for multiple products is the potential cost advantages of producing many closely related goods. A production process exhibits **economies of scope** when the cost of producing multiple goods is less than the aggregate cost of producing each item separately. A convenient measure of such economies is Here, C(Q_1, Q_2) denotes the firm's cost of jointly producing the goods in the respective quantities; C(Q_1) denotes the cost of producing good 1 alone and similarly for C(Q_2). For instance, suppose producing the goods separately means incurring costs of $12 million and $8 million, respectively. The total cost SC _ (Q_1) _ C(Q_2) _ C(Q_1, Q_2) C(Q_1) _ C(Q_2). of jointly producing the goods in the same quantities is $17 million. It follows that SC _ (12 _ 8 _ 17)/(12 _ 8) _ .15. Joint production implies a 15 percent cost savings vis-à-vis separate production. There are many sources for economies of scope. In some cases, a single production process yields multiple outputs. Cattle producers sell both beef and hides; indeed, producing cattle for beef or hides alone probably is not profitable. In other cases, production of a principal good is accompanied by the generation of unavoidable by-products. Often these by-products can be fashioned into marketable products. Sawdust is a valuable by-product of lumber production. Tiny plastic pellets (the by-product of stamping out buttons) are used in sandblasting instead of sand. After the harvest, leftover cornstalks are used to produce alcohol for power generation. Still another source of economies is underutilization of inputs. An airline that carries passengers may find itself with unused cargo space; thus, it contracts to carry cargo as well as passengers. In recent years, many public school systems have made their classrooms available after hours for day-care, after-school, and community programs.

Recent research points to a number of possible reasons. First, the large multiproduct firm is understandably reluctant to risk cannibalizing its existing products by embracing and pursing promising but risky innovations. Second, behavioral factors can play a role—top management is psychologically invested in its current initiatives and consciously or unconsciously embraces the status quo. Finally, diseconomies of scale and scope may play a factor. At large pharmaceutical firms, the high levels of bureaucracy and internal red tape have been blamed for the declining rate of new drug discoveries during the last decade. Attempting to buck this trend, the drug company GlaxoSmithKline has carved dozens of small research units out of its thousand-strong R&D force—each small unit focusing on a single research initiative, with substantial freedom and monetary incentives to succeed. In attempting to emulate the success of biotech firms in basic research, smaller may be better. In turn, Microsoft arguably was held back by diseconomies of scope in extending its operations to browsers and Internet-based computing. Its reputation and inclination for controlling propriety standards made it very difficult to adopt open architectures needed to promote these new operating realms. It would have been better served if it had invested in an independent, stand-alone entity to pursue the browser and Internet-based software markets. Many

experts argue that relying on economies of scale—producing dedicated systems that are economical but inflexible—is no longer enough. The most successful firms in the future will also exploit the flexibility provided by economies of scope.

COST ANALYSIS AND OPTIMAL DECISIONS

Knowledge of the firm's relevant costs is essential for determining sound managerial decisions. First, we consider decisions concerning a single product; then we examine decisions for multiproduct firms.

A Single Product

The profit-maximizing rule for a single-product firm is straightforward: As long as it is profitable to produce, the firm sets its optimal output where marginal revenue equals marginal cost. 6.6 shows a single-product firm that faces a downward-sloping demand curve and U-shaped average cost curves. The firm's profit-maximizing output is Q* (where the MR and MC curves cross), and its optimal price is P* (read off the demand curve). The firm's economic profit is measured by the area of the shaded rectangle in the. The rectangle's height represents the firm's profit per unit (P* _ AC), and its base is total output Q*. (Remember that the firm's average cost includes a normal return on its invested capital. Therefore, a positive economic profit means that the firm is earning a greater-than-normal rate of return.) No alternative output and price could generate a greater economic profit.

By now, the application of marginal revenue and marginal cost should be very familiar. Nonetheless, it is worth pointing out two fallacies that occasionally find their way into managerial discussions. The first fallacy states that the firm always can increase its profit by exploiting economies of scale. But fully exploiting these economies means producing at minimum efficient scale—the point of minimum average cost. 6.6 shows the problem with this contention: The profit-maximizing output Q* falls well short of Q_{min}. In fact, if the firm were to produce at Q_{min}, it would suffer an economic loss. (The demand line falls below the average-cost curve at Q_{min}.) The general point is that the firm's optimal output depends on demand as well as cost. In 6.6, the level of demand for the firm's product is insufficient to justify exploiting all economies of scale. However, we easily could depict a much higher level of demand—one that pushes the firm to an output well above Q_{min}, that is, into the range of increasing average cost. The shows part of a (hypothetical) demand curve and the associated marginal revenue curve that intersects marginal cost at output Q_. For this level of demand, Q_ (a quantity much greater than Q_{min}) is the profit-maximizing output. The second fallacy works in the opposite direction of the first. It states that if the current output and price are unsatisfactory, the firm should raise its price to increase profits. The intuitive appeal of this "rule" is obvious. If price is too low relative to average cost, the remedy is to increase price. However, this contention is not necessarily so. In 6.6, raising price is appropriate only if the current price is lower than P* (with output greater than Q*). If price is already greater than P*, further price increases only reduce profits. In fact, the can readily demonstrate the classic fallacy of managing the product out of business. Suppose management makes the mistake of setting its output at Q_. Here the firm's price P_ is slightly

below average cost, so the firm is incurring a loss. As a remedy, the firm raises price. Does this improve profits? No. The increase in price causes a decrease in quantity (which is expected) but also an increase in average cost (perhaps unexpected). At a higher price and lower output, the firm still is generating a loss. If it raises price again, its volume will shrink further and its price still will fail to catch up with its increasing average cost. By using this strategy, the firm quickly would price itself out of the market.

The Shut-Down Rule

Under adverse economic conditions, managers face the decision of whether to cease production of a product altogether, that is, whether to shut down. Although the choice may appear obvious (shut down if the product is generating monetary losses), a correct decision requires a careful weighing of relevant options. These alternatives differ depending on the firm's time horizon. In the short run, many of the firm's inputs are fixed. Suppose the firm is producing a single item that is incurring economic losses—total cost exceeds revenues or, equivalently, average total cost exceeds price. 6.7 displays the situation. At the firm's current output, average cost exceeds price: AC P*; the firm is earning negative economic profit.

REVIEW QUESTION

1. Define Relevant Costs.
2. What is Economic Profit?
3. Discuss about Fixed and Sunk Costs.
4. What is The Cost of Production?
5. Describe Short-Run Versus Long-Run Cost.
6. What is Returns to Scale and Scope?
7. Describe E-Commerce and Cost Economies.
8. Describe Cost Analysis and Optimal Decisions.

FURTHER READINGS

1. Managerial Economics - Arun Kumar, Rachana Sharma
2. Managerial Economics - Thomas J. Webster
3. Managerial Economics - Petersen / Jain
4. Managerial Economics - Yogesh Maheshwari
5. Managerial Economics - E. Narayanan Nadar, S. Vijayan

UNIT-7 PERFECT COMPETITION

CONTENTS

INTRODUCTION

This Unit and the three that follow focus on the spectrum of industry structures. Markets are typically divided into four main categories: perfect competition, monopolistic competition, oligopoly, and pure monopoly. Table 7.1 provides a preview of these different settings by considering two dimensions of competition: the number of competing firms and the extent of entry barriers. At one extreme (the lower right cell of the table) is the case of perfect competition. Such a market is supplied by a large number of competitors. Because each firm claims only a very small market share, none has the power to control price. Rather, price is determined by supply and demand. As important, there are no barriers preventing new firms from entering the market.

At the other extreme (the upper left cell of the table) lies the case of pure monopoly. Here a single firm supplies the market and has no direct competitors. Thus, as we shall see, the monopolist (if not constrained) has the ultimate power to raise prices and maximize its profit. Clearly, prohibitive entry barriers are a precondition for pure monopoly. Such barriers prevent rival firms from entering the market and competing even handedly with the incumbent monopolist. Oligopoly (shown in the second row of Table 7.1) occupies a middle ground between the perfectly competitive and monopolistic extremes. In an oligopoly, a small number of large firms dominate the market. Each firm must anticipate the effect of its rivals' actions on its own profits and attempt to fashion profit-maximizing decisions in response. Again, moderate or high entry barriers are necessary to insulate the oligopolists from would-be entrants.

Finally, monopolistic competition (not shown in the table) shares several of the characteristics of perfect competition: many small firms competing

in the market and an absence of entry barriers. In this sense, it would occupy the same cell as perfect competition. However, whereas perfect competition is characterized by firms producing identical standardized products, monopolistic competition is marked by product differentiation. In short, the two dimensions of competition shown in Table 7.1, though useful, do not do the full job in distinguishing different market structures.

THE BASICS OF SUPPLY AND DEMAND

A thorough knowledge of the workings of supply and demand, and how they affect price and output in competitive markets, is essential for sound managerial decision making. For example, if a product or service is sold in a perfectly competitive industry, top management is naturally concerned with a prediction of future prices in the market. Should the firm expand capacity with the expectation of price increases? Conversely, if price declines are expected, downsizing might be the proper response.

In a perfectly competitive market, price is determined by the market demand and supply curves. We will consider each of these entities in turn.

The **demand curve** for a good or service shows the total quantities that consumers are willing and able to purchase at various prices, other factors held constant.2 7.1 depicts a hypothetical demand curve D for shoes in a local market. As expected, the curve slopes downward to the right. Any change in price represents a movement along the demand curve.

The **supply curve** for a good or service shows the total quantities that producers are willing and able to supply at various prices, other factors held constant. In 7.1, the supply curve for shoes (denoted by S) is upward sloping. As the price of shoes increases, firms are willing to produce greater quantities because of the greater profit available at the higher price. Any change in price represents a movement along the supply curve.

The **equilibrium price** in the market is determined at point E where market supply equals market demand. 7.1 shows the equilibrium price to be $25 per pair of shoes, the price at which the demand and supply curves intersect. At the $25 price, the quantity of output demanded by consumers exactly matches the amount of output willingly offered by producers. The corresponding equilibrium quantity is 8,000 pairs. To see what lies behind the notion of demand-supply equilibrium, consider the situation at different prices. Suppose the market price were temporarily greater than (say, $35). At this higher price, the amount of shoes firms supply would greatly exceed the amount consumers would purchase. Given the surplus of supply relative to demand, producers would be forced to reduce their prices to sell their output. Price reductions would occur until equilibrium was restored at the $25 price. Similarly, if the price were temporarily lower than $25, consumer demand would outstrip the quantity supplied. The result would be upward pressure on price until the equilibrium price was restored. If we augment the demand and supply graph with quantitative estimates of the curves, we can pinpoint equilibrium price and quantity more precisely. Suppose the market demand curve in 7.1 is described by the equation

Shifts in Demand and Supply

Changes in important economic factors can shift the positions of the demand and/or supply curves, causing, in turn, predictable changes in equilibrium price and quantity. For example, suppose the local economy is coming out of a recession and that consumer incomes are rising. As a result, a greater quantity of shoes would be demanded even at an unchanged price. An increase in demand due to any nonprice factor is depicted as a rightward shift in the demand curve. Shifting the entire curve means that we would expect an increase in the quantity demanded at any prevailing price.4 Such a shift is shown in 7.2a. What is the result of the shift in demand? We see from the that the new equilibrium occurs at a higher price and greater quantity of output. This is hardly surprising. The increase in demand causes price to be bid up. In the process, the amount supplied by firms also increases. The change from the old to the new market equilibrium represents a movement along the stationary supply curve (caused by a shift in demand). Now consider economic conditions that might shift the position of the supply curve. Two principal factors are changes in input prices and technology improvements. For instance, increases in input prices will cause the supply curve to shift upward and to the left. (Any effect that increases the marginal cost of production means that the firm must receive a higher price to be induced to supply a given level of output.) Technological improvements, however, allow firms to reduce their unit costs of production. As a consequence, the supply curve shifts down and to the right. Such a shift is shown in 7.2b. The result is a greater market output and a lower price. The favorable shift in supply has moved the equilibrium toward lower prices and greater quantities along the unchanged demand curve.

COMPETITIVE EQUILIBRIUM

Perfect competition is commonly characterized by four conditions.

1. A large number of firms supply a good or service for a market consisting of a large number of consumers.
2. There are no barriers with respect to new firms entering the market. As a result, the typical competitive firm will earn a zero economic profit.
3. All firms produce and sell identical standardized products. Therefore, firms compete only with respect to price. In addition, all consumers have perfect information about competing prices. Thus, all goods must sell at a single market price.
4. Firms and consumers are price takers. Each firm sells a small share of total industry output, and, therefore, its actions have no impact on price. Each firm takes the price as given—indeed, determined by supply and demand. Similarly, each consumer is a price taker, having no influence on the market price.

It is important to remember that these conditions characterize an ideal model of perfect competition. Some competitive markets in the real world meet the letter of all four conditions. Many other real-world markets are effectively perfectly competitive because they approximate these conditions. At present, we will use the ideal model to make precise price and output predictions for perfectly competitive markets. Later in this and the following Units, we will compare the model to real-world

markets. In exploring the model of perfect competition, we first focus on the individual decision problem the typical firm faces. Then we show how firm-level decisions influence total industry output and price.

Decisions of the Competitive Firm

The key feature of the perfectly competitive firm is that it is a **price taker**; that is, it has no influence on market price. Two key conditions are necessary for price taking. First, the competitive market is composed of a large number of sellers (and buyers), each of which is small relative to the total market. Second, the firms' outputs are perfect substitutes for one another; that is, each firm's output is perceived to be indistinguishable from any other's. Perfect substitutability usually requires that all firms produce a standard, homogeneous, undifferentiated product, and that buyers have perfect information about cost, price, and quality of competing goods. Together, these two conditions ensure that the firm's demand curve is perfectly (or infinitely) elastic. In other words, it is horizontal like the solid price line in 7.3a. Recall the meaning of perfectly elastic demand. The firm can sell as much or as little output as it likes along the horizontal price line ($8 in the). If it raises its price above $8 (even by a nickel), its sales go to zero. Consumers instead will purchase the good (a perfect substitute) from a competitor at the market price. When all firms' outputs are perfect substitutes, the "law of one price" holds: All market transactions take place at a single price. Thus, each firm faces the same horizontal demand curve given by the prevailing market price.

THE FIRM'S SUPPLY CURVE

Part (a) of 7.3 also is useful in describing the supply of output by the perfectly competitive firm. The cost characteristics of the typical firm in the competitive market are as shown in the . The firm faces a U-shaped, average cost curve (AC) and an increasing marginal cost curve (MC). (Recall that increasing marginal cost reflects diminishing marginal returns.)

Suppose the firm faces a market price of $8. (For the moment, we are not saying how this market price might have been established.) What is its optimal level of output? As always, the firm maximizes profit by applying the MR _ MC rule. In the case of perfectly elastic demand, the firm's marginal revenue from selling an extra unit is simply the price it receives for the unit: MR _ P. Here the marginal revenue line and price line coincide. Thus, we have the following rule:

A firm in a perfectly competitive market maximizes profit by producing up to an output such that its marginal cost equals the market price. In 7.3, the intersection of the horizontal price line and the rising marginal cost curve (where P _ MC) identifies the firm's optimal output. At an $8 market price, the firm's optimal output is 6,000 units. (Check for yourself that the firm would sacrifice potential profit if it deviated from this output, by producing either slightly more or slightly less.) Notice that if the price rises above $8, the firm profitably can increase its output; the new optimal output lies at a higher point along the MC curve. A lower price implies a fall in the firm's optimal output. (Recall, however, that if price falls below average variable cost, the firm will produce nothing.) By varying price, we read the firm's optimal output off the

marginal cost curve. The firm's **supply curve** is simply the portion of the MC curve lying above average variable cost.

LONG-RUN EQUILIBRIUM

Perfectly competitive markets exhibit a third important condition: In the long run, firms can freely enter or exit the market. In light of this fact, it is important to recognize that the profit opportunity shown in 7.3a is temporary. Here the typical firm is earning a positive economic profit that comes to __($8.00 _ $6.50)(6,000) _ $9,000. But the existence of positive economic profit will attract new suppliers into the industry, and as new firms enter and produce output, the current market price will be bid down. The competitive price will fall to the point where all economic profits are eliminated.

Market Equilibrium

Let's shift from the typical firm's point of view to that of the market as a whole. 7.4 provides this marketwide perspective. The current equilibrium occurs at E, where the market price is $6 per unit (as in 7.3b) and the industry's total quantity of output is 200,000 units. This output is supplied by exactly 40 competitive firms, each producing 5,000 units (each firm's point of P _ MR _ LMC _ min LAC. minimum LAC). The market is in equilibrium. Industry demand exactly matches industry supply. All firms make zero economic profits; no firm has an incentive to alter its output. Furthermore, no firm has an incentive to enter or exit the industry.

LONG-RUN MARKET SUPPLY

The horizontal line in represents the case of a constant-cost industry. For such an industry, the long-run market supply curve is a horizontal line at a level equal to the minimum LAC of production. Recall that any long-run additions to supply are furnished by the entry of new firms. Furthermore, in a constant-cost industry, the inputs needed to produce the increased industry output can be obtained without bidding up their prices. This is the case if the industry in question draws its resources from large, well-developed input markets. (If the industry is a "small player" in these input markets, an increase in its demand will have a negligible effect on the inputs' market prices.) For instance, the market for new housing exhibits a nearly horizontal long-run supply curve. In the long run, the industry's two main inputs

— building materials and construction labor
— are relatively abundant and provided

For an increasing-cost industry, output expansion causes increases in the price of key inputs, thus raising minimum average costs. Here the industry relies on inputs in limited supply: land, skilled labor, and sophisticated capital equipment. For instance, if U.S. drilling activity increased by 30 percent (perhaps due to increases in world oil prices), the typical oil company's average cost per barrel of oil could be expected to rise, for a number of reasons. First, the increase in drilling would bid up the price of drilling rigs and sophisticated seismic equipment. Second, skilled labor (such as chemical engineering graduates), being in greater demand, would receive higher wages. Third, because the most promising sites are limited, oil companies would resort to drilling marginal sites,

yielding less oil on average. For an increasing-cost industry, the result of such increases in average costs is an upward-sloping long-run supply curve.

MARKET EFFICIENCY

You might be familiar with one of the most famous statements in economics— Adam Smith's notion of an "invisible hand": Every individual endeavors to employ his capital so that its produce may be of greatest value. He generally neither intends to promote the public interest, nor knows how much he is promoting it. He intends only his own security, only his gain. And he is in this led by an invisible hand to promote an end which was no part of his intention. By pursuing his own interest he frequently promotes that of society more effectively than when he really intends to promote it.

One of the main accomplishments of modern economics has been to examine carefully the circumstances in which the profit incentive, as mediated by competitive markets, promotes social welfare.8 although economists are fond of proving theorems on this subject, the present approach is more pragmatic. Our aim is to examine the following proposition: course, getting to the heart of market efficiency requires a careful explanation of what the "efficient" amount of a good or service means.

Private Markets: Benefits and Costs

The main step in our examination of market efficiency is the valuation (in dollar terms) of benefits and costs. We begin the analysis with a single transaction and move on to the thousands of transactions that take place within markets. Consider the following example.

THE DEMAND AND SUPPLY OF DAY CARE A couple is seeking to obtain up to 10 hours of day care per week for their 2-year-old. Through informal inquiries in their neighborhood, they have found a grandmother who has done baby-sitting and some day care in the past and comes highly recommended. The grandmother is not sure whether she is willing to commit to 10 hours. Before any discussion of price takes place, the couple has thought hard about their value for day care. They have decided that the maximum amount they are willing to pay is $8 per hour (that is, they would be indifferent to the options of getting day care at this price and not getting it at all). For her part, the grandmother has decided that her minimum acceptable price is $4. (Thus, $4 is the best estimate of her "cost" based on the value of her time and the strain of taking care of a 2-year-old. All things considered, she just breaks even at this price.) Can the couple and the grandmother conclude a mutually beneficial agreement? How can we measure the parties' gains from an agreement?

The answer to the first question clearly is yes. Any negotiated price between$4 and $8 would be mutually beneficial. What about the second question? If the parties are equally matched bargainers, we might expect the final price to be $6. The grandmother makes a profit of $2 per hour, or $20 per week. Similarly, the couple makes a $2-per-hour "profit"; that is, they pay only $6 for a day-carehour that is worth $8 to them. Their

"profit" per week is $20. The couple's gain (or any consumer's gain in general) is customarily labeled **consumer surplus**.

Although it goes under a different name, the couple's gain is identical in kind (and here in amount) to the grandmother's profit.

THE DAY-CARE MARKET

Let's now extend the previous analysis to the large day-care market that emerged in the last 25 years. 7.6 shows the weekly demand curve for day care in a given geographical region. There is nothing remarkable about this bare-bones demand curve. Depending on the going hourly price for day care, more or less millions of day-care hours will be demanded. The lower the price, the greater the number of hours purchased. However, one aspect of this demand curve (or any demand curve) is important: Besides showing the quantity consumed at any price, the demand curve shows the monetary value that consumers are willing to pay for each unit. For instance, the "first" few units consumed are valued at roughly $12, the demand curve's price intercept. Even at a rate this high, some parents (with high incomes, rotten kids, or both) are willing to pay the high price for day care. But what about the 8 millionth hour of day care consumed? For this hour to be purchased, the hourly price must drop to $4. Put simply, the value of any unit of day care is given by the price the consumer is willing to pay for it.9 (Thus, it is hard to claim that the 8 millionth hour is worth $4.50 because the would-be consumer of this hour is unwilling to pay that high a price.) In short, the value of a particular unit is given by the height of the demand curve at that quantity.10 For this reason, the demand curve can be thought of as a **marginal benefit curve**.

EFFICIENCY AND EQUITY

It is important to emphasize that efficient markets are not necessarily equitable or fair. The outcomes of competitive markets directly reflect the distribution of incomes of those who buy and sell in these markets. An inability to pay excludes many people from the economic equation. In trying to solve the problems of poverty, malnutrition, inadequate health care, and the like, the government has the responsibility of addressing equity issues (as well as efficiency issues).

DYNAMIC, MARKETWIDE EFFICIENCY

In our examination of competitive efficiency, we have focused on a single market and found that the efficient level of output occurs at the intersection of demand and supply, where PC _ MB _ MC. Can this "invisible hand" result be extended to encompass at once all the innumerable markets in a modern economy? The generalization to multiple markets is more complicated than it might seem at first. When dealing with many markets, it is not quite correct to focus on them separately, one at a time. After all, demands for different goods and services in the economy are interdependent. Changing the price of one good affects not only its consumption but also the consumption of substitute and complementary goods. Similarly, any change in price and output in one market generates marginal benefits and costs not only for that good but also for other affected markets. Given these

interdependencies, can we draw any conclusions about the workings of private markets and economic efficiency?

INTERNATIONAL TRADE

As noted in Unit 6, international trade is based on mutually beneficial specialization among countries, that is, on comparative advantage. The final section of this Unit underscores two additional points. First, when free trade is the norm, patterns of trade follow the rules of worldwide supply and demand. If a country's demand outstrips its available supply, it will make up the difference via imports from the rest of the world. Second, the proposition that competitive markets are efficient applies not only to individual markets within a nation but also to all global markets. Free trade is the basis for worldwide efficient production. When nations erect trade barriers, economic welfare is diminished. To see why perfectly competitive global markets are efficient, we use exactly the same arguments as before. Under free trade, firms from all over the world compete for sales to consumers of different nations. Free competition means that the good in question will sell at a single world price (net of transport costs). Only the most efficient lowest-cost firms will supply the good. Only consumers willing and able to pay the world price will purchase the good. Finally, exactly the right amount of the good will be supplied and consumed worldwide. In competitive equilibrium, global output occurs at a quantity such that P _ MB _ MC. The quantity of output is efficient. In a nutshell, this is the efficiency argument for free trade.

Tariffs and Quotas

In reality, worldwide trade is far from free. Traditionally, nations have erected trade barriers to limit the quantities of imports from other countries. Most commonly, these import restrictions have taken the form of tariffs, that is, taxes on foreign goods, or direct quotas. The usual rationale for this is to protect particular industries and their workers from foreign competition. Since World War II, the industrialized nations of the world have pushed for reductions in all kinds of trade barriers. Under the General Agreement on Tariffs and Trade (GATT), member nations meet periodically to negotiate reciprocal cuts in tariffs. In the last decade, there has been a rise in protectionist sentiment in the United States, aimed in part at insulating domestic industries from competition and, in part, as retaliation against alleged protectionist policies by Japan and Europe. Although there are a number of strategic reasons why a country might hope to profit from trade barriers, the larger problem is the efficiency harm imposed by these restrictions. To illustrate this point, we return to the digital watch example introduced in Unit 6.

REVIEW QUESTION

1. What are the basics of supply and demand?
2. Describe shifts in demand and supply.
3. Describe competitive equilibrium also decisions of the competitive firm.
4. What is the firm's supply curve?
5. Describe long-run equilibrium. What is long-run market supply?

6. Describe market efficiency. What is the demand and supply of day care?
7. What is the efficiency and equity? Describe dynamic, marketwise efficiency
8. Write a short note on international trade.

FURTHER READINGS

1. Managerial Economics - Arun Kumar, Rachana Sharma
2. Managerial Economics - Thomas J. Webster
3. Managerial Economics - Petersen / Jain
4. Managerial Economics - Yogesh Maheshwari
5. Managerial Economics - E. Narayanan Nadar, S. Vijayan

UNIT-8 MONOPOLY

CONTENTS

INTRODUCTION

Let's start by considering a monopolist's price and output decision. Being the lone producer, the monopolist is free to raise price without worrying about losing sales to a competitor that might charge a lower price. Although the monopolist has complete control over industry output, this does not mean it can raise price indefinitely. Its optimal price and output policy depends on market demand. Because the monopolist is the industry, its demand curve is given simply by the industry demand curve. 8.1 depicts the industry demand curve and long-run costs for the monopolist. Given information on demand and cost, it is straightforward to predict monopoly price and output. As a profit maximizer, the monopolist should set its output such that marginal revenue (derived from the industry demand curve) equals the marginal cost of production. In the , this output, Q_M, is shown where the monopolist's marginal revenue and marginal cost curves intersect. According to the industry demand curve, the corresponding monopoly price is P_M. The area of the shaded rectangle measures the monopolist's total excess profit. This profit is the product of the monopolist's profit per unit, P_M _ AC (the rectangle's height), and total output, Q_M (the rectangle's base).

We should make two related remarks about the potential for excess profits under pure monopoly. First, monopoly confers a greater profit to the firm than it would have if the firm shared the market with competitors. We have seen that economic profits in perfect competition are zero in the long run—not so for the monopolist. Second, even when the firm occupies a pure-monopoly position, its excess profits depend directly on the position of industry demand versus its cost. 8.2 makes the point by depicting three different industry demand curves. It should be evident that only curve D_1 offers significant excess profits. Demand curves D_2 and D_3 offer very little in the way of profit possibilities. Although they differ with respect to elasticities, both curves barely exceed the monopolist's average cost. The lesson here is that pure monopoly enables the firm to earn excess profit, but the actual size of this profit depends on a comparison of demand and cost. For instance, if other goods or services are close substitutes for the monopolist's product, industry demand may be relatively elastic and afford relatively little excess profit (curve D_2). If it is to increase its profit substantially, the monopoly firm must find a way to lower its average cost of production or to raise market demand. (However, there may be no demand at all for the

monopolist's unique product. The U.S. Patent Office overflows with inventions that have never earned a dime.)

Barriers to Entry

A **barrier** is any factor that blocks or impedes entry of new firms into a particular market. There is a wide variety of barriers to entry that are more or less important, depending on the market under consideration. In some cases, one or more of these barriers are sufficient to support a single dominant firm in the market. In others, entry barriers are not absolute but limit the market to a small number of firms. It is also useful to speak of barriers to competition—that is, factors that, while not precluding rivals from the market, insulate a given firm from direct competition. Sources of entry barriers include the following.

ECONOMIES OF SCALE When average cost falls significantly with increases in scale, a new firm must enter the market with a large market share to be competitive. If this addition to industry output requires a significant drop in market price, entry will be unprofitable. In so-called natural monopolies, averagecost continually decreases with output, implying that a single firm achieves the lowest possible unit cost by supplying the entire market. For instance, it is cheaper for one company to lay a single network of cables to provide cable TV to a particular town or region.

CAPITAL REQUIREMENTS In some industries (automobiles, defense, oil refining, deep-sea drilling), the capital requirements of production are enormous. In others (chemicals, pharmaceuticals, electronics), large investments in research and development are necessary. When large sunk costs are required, entry is particularly risky. (If, after entry, a firm finds itself suffering losses, it will be largely unable to recover its investment.)

PURE QUALITY AND COST ADVANTAGES Sometimes a single firm has absolute quality or cost advantages over all potential competitors. Cost advantages may be due to superior technology, more efficient management, economies of scope, or learning. For these reasons, Intel dominates the market for microchips, Wal-Mart is the world's leading chain of discount department stores, and Boeing and Airbus share the global aircraft market. In many e-commerce markets, network externalities (making larger networks more valuable to customers) bestow an important quality advantage on the market leader (eBay in online auctions for instance). Although there are many close substitutes, Coca-Cola continues to guard the secret for its best-selling soft drink. In the 1980s and 1990s, the Department of Defense used sole-source procurements to purchase major weapon systems, claiming that only a single qualified supplier existed. A dramatic expression of the monetary return to "being the best" is the annual income of a "superstar" such as Tiger Woods, Kobe Bryant, Lady Gaga, George Clooney, or Angelina Jolie.

PRODUCT DIFFERENTIATION Once an incumbent has created a preference for a unique product or brand name via advertising and marketing campaigns, it has erected considerable barriers to new entrants that seek to compete for its customers. Producers of retail goods and services thrive on product differentiation, real or perceived.

Differentiation is the norm in products ranging from soft drinks to ready-to-eat breakfast cereals to toothpaste. Switching costs can be an important barrier to competition in markets for information-intensive goods and services. When customers have invested in learning to use a particular software program, navigate a Web site, or set up online accounts, they are less likely to switch to competitive (perhaps even superior) alternatives. Google's continuing dominance in Internet search depends in part on the high learning costs of changing to an alternative search engine.

CONTROL OF RESOURCES A barrier to entry exists when an incumbent firm (or firms) controls crucial resources—mineral deposits, oil supplies, even scientific talent. At the local level, a retailer's choice location may provide protection from entry by would-be competitors. Ownership of unique items (fine art, antiques) confers a degree of monopoly power (albeit limited by the availability of substitutes). For instance, the price of a unique item at auction is determined by what the market will bear, not by competitive supply. The best-known examples of monopoly power based on resource control include French champagne, De Beers (diamonds), and OPEC (crude oil).

PATENTS, COPYRIGHTS, AND OTHER LEGAL BARRIERS A patent grants the holder exclusive rights to make, use, or sell an invention for 20 years. A patent can apply to an idea, process, or system as well as to an invention. A copyright prohibits the unauthorized copying of a particular work. (Currently, there is considerable controversy concerning whether computer software qualifies for copyright protection.) Patents and copyrights constitute important barriers to entry in computers, machinery, electronics, publishing, pharmaceuticals, defense, and chemicals. In many instances (local utilities, cable television firms, vendors on state highways and in national parks), the government grants legal monopolies for extended periods of time.

STRATEGIC BARRIERS Finally, the dominant firm (or firms) may take actions explicitly aimed at erecting entry barriers. Securing legal protection (via patent or copyright) is only one example. A monopolist may exercise limit pricing, that is, keep price below monopoly levels to discourage new entry. It may threaten retaliatory pricing. For the same reasons, it may engage in extensive advertising and brand proliferation, not because this is profitable in itself (it may not be) but to raise the cost of entry for new competitors. Finally, the monopolist may intentionally create excess productive capacity as a warning that it can quickly expand output should a new firm attempt to enter. We will re-examine strategic barriers in Unit 10.

Intel Corporation is by far the most powerful and profitable producer of microchips in the world. In the early 1970s, Intel invented the microprocessor, the computer on a chip that serves as the "brain" of the personal computer. Since then, it has produced numerous generations of chips, including the Pentium series and more recently the Itanium series, each faster and cheaper than the last. At the close of 2010, it accounted for 81 percent of the world's semiconductor market, a share mainly unchanged over the past decades. In advanced microprocessors, its

market dominance is well over 90 percent. Thus, Intel has held a virtual monopoly in the microchip market.1 Over the years, however, new competitors have increasingly pushed into Intel's markets. In the mid-1990s, other chips emerged as competitors in particular market segments: the Power PC chip shared by IBM, Motorola, and
Apple, Hewlett-Packard's RISC chip, and Sun's SPARC chip, to name a few. 324

PERFECT COMPETITION VERSUS PURE MONOPOLY

Recall from Unit 7 that a perfectly competitive market delivers output to consumers at the lowest sustainable price. (If prevailing prices were any lower, firms would incur losses and leave the market.) In a pure monopoly, in contrast, a single firm is the sole supplier of a good or service. The monopolist uses its market power to restrict output and raise price.

The simplest way to compare and contrast the basic price and output implications for purely monopolistic and purely competitive industries is by means of a graph. 8.3 displays demand and cost curves for an unspecified good or service. The industry demand curve D has the usual downward slope. For any given industry price, it predicts total industry-wide sales. The horizontal cost line S depicts the long-run unit cost of supplying different industry levels of output. The cost line reflects the fact that output can be expanded in the long run at a constant cost (at least for the range of output shown in the graph). We can now use these demand and cost facts to predict long-run price and output for a perfectly competitive industry versus the same industry organized as a pure monopoly.

Under perfect competition, industry price and output are determined at the intersection of the demand and supply curves. The total industry output is split among a large number of firms, each producing at a constant cost per unit. Competitive price and output are PC and QC, respectively. Note that PC is identical to the typical supplier's cost per unit; that is, the typical competitive firm makes zero economic profit. If the market price ever rises above unit cost, opportunities for positive economic profits will induce suppliers, including new entrants, to increase output. This supply influx will drive price back down to the unit cost level.2

Now suppose the same industry is controlled by a single firm, a monopolist. Because the monopolist is the industry, its demand curve is simply D. The monopolist can supply as much or as little output as it wishes at a constant unit cost given by S. What price and output will a profit-maximizing monopolist set? As always, marginal analysis supplies the answer: The firm will set output where industry-wide marginal revenue equals marginal cost. 8.3 shows MR (derived from the industry demand curve in the usual way). The line S does double duty: Besides being a supply curve, it measures the monopolist's marginal cost curve. (The monopolist can produce additional units at this unit cost.) The monopolist's optimal output is QM (where MR _ MC), and the required market-clearing price is PM.

8.3 provide a graphical comparison of perfect competition and pure monopoly. Under competition, long-run price is driven down to the

lowest sustainable level (where industry economic profit is zero). As a consequence, a competitive market delivers maximum benefits to consumers. In contrast, the monopolist has the opportunity to exercise market power, that is, to raise price above competitive levels. The monopolist does not set price and output capriciously. The key to maximizing monopoly profit is to restrict output to well below the competitive level and, in so doing, to raise price. The monopolist's optimal level of output occurs where marginal revenue equals marginal cost. Note that monopoly output is always smaller than competitive output. In the, the intersection of MR and MC occurs to the left of the intersection of D and MC. Thus, we have the following summary comparison of perfect competition and pure monopoly: and Competition delivers output at a minimum price and implies zero industry profits. Monopoly delivers maximum industry profits by limiting output and raising price.

Finally, the presence of monopoly represents a major deviation from the efficiency of perfect competition. In 8.3, the net benefit attained under perfect competition is measured by the area of the large consumer-surplus triangle ACE. (Producers make zero economic profits because PC _ AC in the long run.) By contrast, under pure monopoly, the monopolist raises price, thereby earning a pure economic profit (rectangle BCDM) but leaving a smaller triangle of surplus for the consumer (triangle ABM). Thus, under monopoly, the sum of consumer surplus and producer profit is given by the trapezoidal area ACDM, which is smaller than the total gains under perfect competition by the triangle MDE.

The triangle MDE is referred to as the **deadweight loss** attributed to monopoly. The economic critique of monopoly is not simply that the firm gains at the expense of consumers when it elevates price. (In terms of total welfare, the firm's profit counts equally with the consumers' surplus. Indeed, consumers could well be shareholders of the monopolist and share in the profit directly.) Rather, the important point is that the monopolist's elevation of price and restriction of output cause a reduction in total welfare. The reduction in consumer surplus (relative to the competitive outcome) exceeds the excess profit earned by the monopolist. The deadweight-loss triangle (MDE) measures the size of the total welfare loss.

Put another way, this deadweight loss would be regained if market output were increased from QM to QC. For these additional units, consumers' marginal benefits exceed suppliers' marginal costs. Consequently, producing this output would increase social welfare. As we will see later in this Unit, the common government response to the so-called case of "natural" monopoly is to regulate lower prices and increased output. Similarly, as will be noted in Unit 11, the government undertakes a broad spectrum of antitrust initiatives to restrain or prohibit specific actions and behavior that would lead to monopolization of markets.

Cartels

A **cartel** is a group of producers that enter into a collusive agreement aimed at controlling price and output in a market. The intent of the cartel

is to secure monopoly profits for its members. Successful maintenance of the cartel not only has an immediate profit advantage; it also reduces the competitive uncertainties for the firms and can raise additional entry barriers to new competitors. In the United States, collusive agreements among producers (whether open or tacit) represent violations of antitrust laws and are illegal.3 Some cartels outside the United States have the sanction of their host governments; in others, countries participate directly. The best-known and most powerful cartels are based on control of natural resources. In the 1990s and today, the Organization of Oil Exporting Countries (OPEC) controls about 40 percent of the world supply of oil. De Beers currently controls the sale of more than 90 percent of the world's gem-quality diamonds. The monopoly model is the basis for understanding cartel behavior. The cartel's goal is to maximize its members' collective profit by acting as a single monopolist would. Based on the demand it faces, the cartel maximizes profit by restricting output and raising price. Ideally, the cartel establishes total output where the cartel's marginal revenue equals its marginal cost. For instance, if cartel members share constant and identical (average and marginal) costs of production, 8.3's depiction of the monopoly outcome would apply equally to the cartel. The cartel maximizes its members' total profits by restricting output and raising price according to Q_M and P_M, where marginal revenue equals marginal cost.4 Output restriction is essential for a cartel to be successful in maximizing its members' profits. No matter how firm its control over a market, a cartel is not exempt from the law of demand. To maintain a targeted price, the cartel must carefully limit the total output it sells. Efforts to sell additional output lead to erosion of the cartel price. The larger the additions to supply, the greater the fall in price and, therefore, the greater the decline in the cartel's total profit. This observation underscores the major problem cartels face: Cartels are inherently unstable. The reason lies in the basic conflict between behavior that maximizes the collective profits of the cartel and self-interested behavior by individual cartel members.

To see this, return to the cartel's optimal price and output, P_M and Q_M, in 8.3. Suppose the cartel agrees to set total output at Q_M and assigns production quotas to members. The self-interest of each member is to overproduce its quota. The member can sell this additional output by cutting price very slightly. (Remember that one member's additional output is small enough t o put little downward pressure on price.) What effect does this added output have on the member's profit? 8.3 shows that the cartel price is well above marginal cost. Thus, even allowing for a slightly discounted selling price, sellingthe extra output is very profitable. Each member has an incentive to cheat on its agreed-upon output quota. But if all members overproduce, this behaviour is self-defeating. If all members increase output (say, by 10 to 15 percent),flooding the market with extra output will have a significant downward effect on price. The total output of the cartel will be far greater than Q_M, price will fall below P_M, and the cartel's profit inevitably must drop. Thus, overproduction is a constant threat to the cartel's existence.5 In the presence of wholesale cheating, the cartel may fall apart.

The 11 member nations of OPEC meet twice a year to discuss the cartel's target price for crude oil and to allot members' production quotas. Like a continuing drama with many acts, the OPEC negotiations center on (1) an assessment of the world demand for oil, (2) the appropriate limit on total OPEC supply, and (3) the division of this supply among cartel members. Over the last 15 years, OPEC has had a mixed record in limiting its supply and maintaining high oil prices.6 Until mid-2001, OPEC was largely successful in negotiating lower total output levels for the cartel and, therefore, maintaining high crude oil prices. OPEC successively cut its total output quota from 26 million to 24.2 million barrels per day (mbd), members largely honoured their individual quotas, and prices rose to above $40 per barrel. However, with the worldwide economic slowdown in 2002 and greatly increased supply by non member Russia, OPEC faced the prospect of soft and falling oil prices. With OPEC members exceeding their quotas by an estimated 1 million total barrels per day, oil prices fell below $20 per barrel.

Natural Monopolies

A **natural monopoly** occurs when the average cost of production declines throughout the relevant range of product demand. Utilities—water, electric power, gas, and telephone—typically fall into this category. 8.4 shows a natural monopoly (say, in the generation of electricity) that displays steeply declining average cost.

Natural monopoly poses obvious difficulties for the maintenance of workable competition. First, it is costly and inefficient for multiple competing firms to share the market. A single firm can always produce a specified quantity of output—call this Q—at lower average cost than it could if the same total quantity were supplied by n firms, each producing Q/n. (Use 8.4 to confirm this.) For six local firms to make the large capital investment to supply electricity is unnecessarily duplicative and costly. With a facility of suitable capacity, a single firm is better suited to be the sole source of supply. Second, even if the market, in principle, could support more than one firm, the inevitable result would be the emergence of a single dominant monopolist. This is simply to say that any firm that increases output can achieve lower unit costs and so price the competition out of the market. Thus, we would expect that the first firm to enter the market and expand its output will grow to control the industry.

Government decision makers play an active and direct role in the regulation of natural monopoly. The principal regulatory aim is to target industry price and output at the efficient competitive level. Let's use 8.4 to display the natural-monopoly outcome, with and without price regulation. In the absence of any regulation (i.e., under a policy of laissez-faire), the firm acts as a pure monopolist. The resulting outcome is the price-quantity pair Q_M and P_M, where the firm's marginal revenue equals its marginal cost. Here the marginal benefit of the last unit consumed is equal to the monopoly price, which, of course, is well above the marginal cost of production. An increase in output from the monopoly level would improve welfare (since MB _ MC). The regulator can induce an increase in output by limiting the natural monopolist to a price that delivers a "fair" rate of return on the firm's investment. This is

accomplished by instituting **average-cost pricing**. The appropriate price and quantity are determined by the intersection of the demand and average cost curves in 8.4. At price PR, the firm earns zero "economic" profit; that is, price exactly equals average cost, where AC includes a provision for a normal return on invested capital.

MONOPOLISTIC COMPETITION

In perfect competition, all firms supply an identical standardized product. In monopoly, a single firm sells a unique product (albeit one that may have indirect substitutes). As the term suggests, **monopolistic competition** represents a mixture of these two cases. The main feature of monopolistic competition is product differentiation: Firms compete by selling products that differ slightly from one another. Product differentiation occurs to a greater or lesser degree in most consumer markets. Firms sell goods with different attributes (claimed to be superior to those of competitors). They also deliver varying levels of support and service to customers. Advertising and marketing, aimed at creating product or brand-name allegiance, reinforce (real or perceived) product differences. Product differentiation means that competing firms have some control over price. Because competing products are close substitutes, demand is relatively elastic, but not perfectly elastic as in perfect competition. The firm has some discretion in raising price without losing its entire market to competitors. Conversely, lowering price will induce additional (but not unlimited) sales. In analyzing monopolistic competition, one often speaks of product groups. These are collections of similar products produced by competing firms. For instance, "designer dresses" would be a typical product group, within which there are significant perceived differences among competitors. The determination of appropriate product groups always should be made on the basis of substitutability and relative price effects. Many, if not most, retail stores operate under monopolistic competition. Consider competition among supermarkets. Besides differences in store size, types of products stocked, and service, these stores are distinguished by locational convenience—arguably the most important factor. Owing to locational convenience and other service differences, a spectrum of different prices can persist across supermarkets without inducing enormous sales swings toward lower-priced stores.

Monopolistic competition is characterized by three features. First, firms sell differentiated products. Although these products are close substitutes, each firm has some control over its own price; demand is not perfectly elastic. Second, the product group contains a large number of firms. This number (be it 20 or 100) must be large enough so that each individual firm's actions have negligible effects on the market's average price and total output. In addition, firms act independently; that is, there is no collusion. Third, there is free entry into the market. One observes that the last two conditions are elements drawn from perfect competition. Nonetheless, by virtue of product differentiation (condition 1), the typical firm retains some degree of monopoly power. Let's consider the output and price implications of these conditions. 8.5a shows a short-run equilibrium of a typical firm under monopolistic competition. Because of product differentiation, the firm faces a slightly downward-sloping

demand curve. (If it raises price slightly, it loses some, but not all, customers to competitors.) Given this demand curve, the firm maximizes profit by setting its marginal revenue equal to its marginal cost in the usual way. In the , the resulting output and price are Q and P, respectively. Because price exceeds average cost, this typical firm is earning positive economic profits. In a long-run equilibrium, the free entry (or exit) of firms ensures that all industry participants earn zero economic profits. Thus, in the long run, the outcome in 8.5a is not sustainable. Attracted by positive economic profits, new firms will enter the market. Because it must share the market with a greater number of competitors, the typical firm will find that demand for its product will be reduced; that is, its demand curve will shift to the left. 8.5b shows the firm's new long-run demand curve. As in part (a), the firm is profit maximizing. The firm's optimal output is Q_E, where marginal revenue equals marginal cost. However, even as a profit maximizer, the firm is earning zero economic profit. At this output, its price, P_E, exactly equals its average cost. In fact, the firm's demand curve is tangent to (and otherwise lies below) its average cost curve. Any output other than Q_E, greater or smaller, implies an economic loss for the firm.

REVIEW QUESTION

1. What are the barriers to entry in monopoly market?
2. What is the difference between perfect competition and pure monopoly?
3. Describe natural monopolies.
4. Discuss about monopolistic competition.

FURTHER READINGS

1. Managerial Economics - Arun Kumar, Rachana Sharma
2. Managerial Economics - Thomas J. Webster
3. Managerial Economics - Petersen / Jain
4. Managerial Economics - Yogesh Maheshwari
5. Managerial Economics - E. Narayanan Nadar, S. Vijayan

UNIT-9 OLIGOPOLY

CONTENTS

INTRODUCTION

In the previous two Units, we focused on perfect competition and pure monopoly, the polar cases of market structure. However, many markets occupy positions between these extremes; that is, they are dominated by neither a single firm nor a plethora of firms. Oligopoly is the general category describing markets or industries that consist of a small number of firms.

Because of oligopoly's importance and because no single model captures the many implications of firm behavior within oligopoly, we devote the entire Unit to this topic.

A firm within an oligopoly faces the following basic question: How can it determine a profit-maximizing course of action when it competes against an identifiable number of competitors similar to itself? This Unit and the succeeding Unit on game theory answer this question by introducing and analysing competitive strategies. Thus, we depart from the approach taken previously where the main focus was on "single" firm facing rivals whose actions are predictable and unchanging. In crafting a competitive strategy, a firm's management must anticipate a range of competitor actions and be prepared to respond accordingly. Competitive strategy finds its most important applications within oligopoly settings. By contrast, in a pure monopoly, there are no immediate competitors to worry about. In pure competition, an individual firm's competitive options are strictly limited. Industry price and output are set by supply and demand, and the firm is destined to earn a zero profit in the long run.

The strategic approach extends the single-firm point of view by recognizing that a firm's profit depends not only on the firm's own actions but also on the actions of competitors. Thus, to determine its own optimal action, the firm must correctly anticipate the actions and reactions of its rivals. Roughly speaking, a manager must look at the competitive situation not only from his or her own point of view but also from rivals' perspectives. The manager should put himself or herself in the competitor's place to analyze what that person's optimal decision might be. This approach is central to game theory and is often called interactive or strategic thinking. The outline of this Unit is as follows. In the first section, we describe how to analyze different types of oligopolies, beginning with Michael Porter's Five-Force model. Next, we introduce the concept of market concentration, as well as the link

between concentration and industry prices. In the following section, we consider two kinds of quantity competition: when a market leader faces a number of smaller competitors and when competition is between equally positioned rivals. In the third section, we examine price competition, ranging from a model of stable prices based on kinked demand to a description of price wars. Finally, in the fourth section, we explore two other important dimensions of competition within oligopolies: the effects of advertising and of strategic pre-commitments.

OLIGOPOLY

An **oligopoly** is a market dominated by a small number of firms, whose actions directly affect one another's profits. In this sense, the fates of oligopoly firms are interdependent. To begin, it is useful to size up an oligopolistic industry along a number of important economic dimensions.

Five-Force Framework

For 25 years, Michael Porter's Five-Forces model has provided a powerful synthesis for describing the structures of different industries and guiding competitive strategy.1 9.1 provides a summary of the Five-Forces framework. The core of Porter's analysis centers on internal industry rivalry: the set of major firms competing in the market and how they compete. Naturally, the number of close rivals, their relative size, position, and power, are crucial. (The following section looks closely at the notion of industry concentration to measure the number and sizes of firms.) Entry into the market is the second most important factor in sizing up the industry. We have already seen that freeentry predisposes a perfectly competitive market to zero economic profits in the long run. Conversely, significant barriers to entry (as listed and described in Unit 8) are a precondition for monopoly. Ease of entry is also crucial for analyzing oligopoly. Boeing and Airbus compete vigorously to sell new aircraft, but barriers to entry due to economies of scale protect them from new competitors. By contrast, numerous new discount airlines in the United States and Europe have dramatically changed the competitive landscape in the air travel market. Similarly, a small independent studio (putting together a good script, directing talent, and up-and-coming actors) can produce a well-reviewed and profitable hit movie despite the formidable clout of the major studios. The impacts of substitutes and complements directly affect industry demand, profitability, and competitive strategy. In a host of industries, this impact is ongoing, even relentless. For instance, trucking and railways are substitutes, competing modes of transport in the long-haul market. Soft-drink consumption suffers at the hands of bottled water, sports drinks, and new-age beverages. In other cases, the emerging threat of new substitutes is crucial.

Industry Concentration

As noted earlier, an oligopoly is dominated by a small number of firms. This "small number" is not precisely defined, but it may be as small as two (a duopoly) or as many as eight to ten. One way to grasp the numbers issue is to appeal to the most widely used measure of market structure: the **concentration ratio**. The four-firm concentration ratio is the percentage of sales accounted for by the top four firms in a market or

industry. (Eight-firm and twenty-firm ratios are defined analogously.) Concentration ratios can be computed from publicly available market-share information. Ratios also are compiled in the U.S. Census Bureau, released by the government at five-year intervals. Table 9.1 lists concentration ratios for selected goods and services compiled from both sources. Notice the progression from highly concentrated to less concentrated industries.

Market concentration has a ready interpretation. The higher the concentration ratio, the greater is the degree of market dominance by a small number of firms. Indeed, a common practice is to distinguish among different market structures by degree of concentration. For example, an **effective monopoly** is said to exist when the single-firm concentration ratio is above 90 percent, $CR_1 \_ 90$. A market may be viewed as **effectively competitive** when CR_4 is below 40 percent. If $CR_4 \_ 40$ percent, the top firms have individual market shares averaging less than 10 percent, and they are joined by many firms with still smaller market shares. Finally, one often speaks of a **loose oligopoly** when 40 percent $\_ CR_4 \_ 60$ percent and a **tight oligopoly** when $CR_4 \_ 60$ percent. Monopolistic competition, discussed in the previous Unit, typically falls in the loose-oligopoly range.

About three-quarters of the total dollar value of goods and services (gross domestic product or GDP) produced by the U.S. economy originate in competitive markets, that is, markets for which $CR_4 \_ 40$. Competitive markets included the lion's share (85 percent or more) of agriculture, forestry, fisheries, mining, and wholesale and retail trade. Competition is less prevalent in manufacturing, general services, and construction (making up between 60 and80 percent of these sectors). In contrast, pure monopoly accounts for a small portion of GDP (between 2 and 3 percent). Tight oligopolies account for about 10 percent of GDP, whereas loose oligopolies comprise about 12 percent.[2] In short, as Table 9.1 shows, while concentrated markets are relatively rare in the U.S. economy, specific industries and manufactured products are highly concentrated.

Because the notion of concentration ratio is used so widely, it is important to understand its limitations. The most serious limitation lies in the identification of the relevant market. A market is a collection of buyers and sellers exchanging goods or services that are very close substitutes for one another. (Recall that the cross-elasticity of demand is a direct measure of substitution. The larger the impact on a good's sales from changes in a competitor's price, the stronger the market competition.) Concentration ratios purport to summarize the size distribution of firms for relevant markets. However, it should be evident that market definitions vary, depending on how broadly or narrowly one draws product and geographic boundaries. First, in many cases the market definitions used in government statistics are too broad. An industry grouping such as pharmaceutical products embraces many distinct, individual product markets. Numerous firms make up the overall consumer- drug market (concentration is low), but individual markets (drugs for ulcers and blood pressure) are highly concentrated. Similarly,

government statistics encompass national markets and therefore cannot capture local monopolies.

Concentration and Prices

Concentration is an important factor affecting pricing and profitability within markets. Other things being equal, increases in concentration can be expected to be associated with increased prices and profits. One way to make this point is to appeal to the extreme cases of pure competition and pure monopoly. Under pure competition, market price equals average cost, leaving all firms zero economic profits (i.e., normal rates of return). Low concentration leads to minimum prices and zero profits. Under a pure monopoly, in contrast, a single dominant firm earns maximum excess profit by optimally raising the market price. Given these polar results, it is natural to hypothesize a positive relationship between an industry's degree of monopoly (as measured by concentration) and industry prices. For instance, the smaller the number of firms that dominate a market (the tighter the oligopoly), the greater is the likelihood that firms will avoid cutthroat competition and succeed in maintaining high prices. High prices may be a result of tacit collusion among a small number of equally matched firms. But even without any form of collusion, fewer competitors can lead to higher prices. The models of price leadership and quantity competition (analyzed in the next section) make exactly this point. There is considerable evidence that increases in concentration promote higher prices. The customary approach in this research is to focus on particular markets and collect data on price (the dependent variable) and costs, demand conditions, and concentration (the explanatory variables). Price is viewed in the functional form where C denotes a measure of cost, D a measure of demand, and SC seller concentration. Based on these data, regression techniques are used to estimate this price relationship in the form of an equation. Of particular interest is the separate influence of concentration on price, other things (costs and demand) being equal. The positive association between concentration and price has been confirmed for a wide variety of products, services, and markets—from retail grocery chains to air travel on intercity routes; from cement production to television advertising; from auctions of oil leases and timber rights to interest rates offered by commercial banks. More generally, a large-scale study of manufacturing (using five-digit product categories) forthe 1960s and 1970s shows that concentration has an important effect on prices for consumer goods and materials (and a smaller positive effect for capital and producer goods). Is an increase in monopoly power necessarily harmful to the interests of consumers? The foregoing discussion citing the evidence of higher prices would say yes. However, an alternative point of view claims that monopoly (i.e., large firms) offers significant efficiency advantages vis-á-vis small firms. According to this hypothesis, monopoly reflects superior efficiency in product development, production, distribution, and marketing. A few firms grow large and become dominant because they are efficient. If these cost advantages are large enough, consumers can obtain lower prices from a market dominated by a small number of large firms than from a competitive market of small firms.Thus, a price comparison between a tight

oligopoly and a competitive market depends on which is the greater effect: the oligopoly's cost reductions or its price increases. For example, suppose that in the competitive market P_C _ AC_C, and, in the tight oligopoly P_O _ $1.15AC_O$. Absent a cost advantage, the oligopoly exhibits higher prices. But if the oligopoly's average cost advantage exceeds 15 percent, it will have the lower overall price.

QUANTITY COMPETITION

There is no single ideal model of competition within oligopoly. This is hardly surprising in view of the different numbers of competitors (from two upward) and dimensions of competition (price, product attributes, capacity, technological innovation, marketing, and advertising) encompassed by oligopoly. In this section, we examine quantity competition in a pair of settings. In the following section, we take up different kinds of price competition.

A Dominant Firm

In many oligopolistic industries, one firm possesses a dominant market share and acts as a leader by setting price for the industry. (Price leadership also is possible among equals.) Historically, one can point to dominant firms, such as General Motors in the automobile industry, Du Pont in chemicals, and U.S. Steel. Firms that currently hold dominant market shares include IBM in mainframe computers, eBay in online auctions, Federal Express in overnight delivery, Intel in microchips, and Microsoft in PC software, to name just a few.

Competition among Symmetric Firms

Now let's modify the previous setting by considering an oligopoly consisting of a small number of equally positioned competitors. As before, a small number of firms produce a standardized, undifferentiated product. Thus, all firms are locked into the same price. The total quantity of output supplied by the firms determines the prevailing market price according to an industry demand curve. Via its quantity choice, an individual firm can affect total output and therefore influence market price.

A simple but important model of quantity competition between duopolists (i.e., two firms) was first developed by Augustin Cournot, a nineteenth-century French economist. To this day, the principal models of quantity competition bear his name. Knowing the industry demand curve, each firm must determine the quantity of output to produce—with these decisions made independently. As a profit maximizer, what quantity should each firm choose? To answer this question, let's consider the following example.

PRICE COMPETITION

In this section, we consider two basic models of price competition. The first is a model of stable prices based on kinked demand. The second is a model of price wars based on the paradigm of the prisoner's dilemma.

Price Rigidity and Kinked Demand

Competition within an oligopoly is complicated by the fact that each firm's actions (with respect to output, pricing, advertising, and so on) affect the profitability of its rivals. Thus, actions by one or more firms typically will trigger competitive reactions by others; indeed, these

actions may trigger "second round" actions by the original firms. Where does this jockeying for competitive position settle down? (Or does it settle down?) We begin our discussion of pricing behavior by focusing on a model of stable prices and output. Many oligopolies— steel, automobiles, and cigarettes, to name a few—have enjoyed relatively stable prices over extended periods of time. (Of course, prices adjust over time to reflect general inflation.) Even when a firm's cost or demand fluctuates, it may be reluctant to change prices.

Price rigidity can be explained by the existence of **kinked demand curves** for competing firms. Consider a typical oligopolist that currently is charging price P*. Why might there be a kink in its estimated demand curve, as in 9.3? Suppose the firm lowers its price. If price competition among firms is fierce, such a price cut is likely to be matched by rival firms staunchly defending their market shares. The upshot is that the firm's price reduction will generate only a small increase in its sales. (The firm will not succeed in gaining market share from its rivals, although it could garner a portion of the increase in industry sales owing to lower market wide prices.) In other words, when it comes to price reductions, demand is relatively inelastic. Conversely, suppose the firm raises its price above P*. By holding to their present prices, rival firms can acquire market share from the price raiser. If the other firms do not follow, the firm will find its sales falling precipitously for even small price increases. In short, demand is elastic for price increases. This explains the demand curve's kink at the firm's current price. In view of kinked demand, the firm's profit-maximizing price and quantity are simply P* and Q*. This is confirmed by noting that the firm's marginal revenue curve in 9.3 is discontinuous. The left part of the MR curve corresponds to the demand curve to the left of the kink. But MR drops discontinuously if price falls slightly below P*. The presence of the vertical discontinuity in MR means that P* and Q* are optimal as long as the firm's marginal cost curve crosses MR within the gap. The dotted MC curve in the shows that marginal cost could decrease without changing the firm's optimal price. (Small shifts in demand that retain the kink at P* would also leave the firm's optimal price unchanged.) In short, each firm's price remains constant over a range of changing market conditions. The result is stable industry-wide prices.

The kinked demand curve model presumes that the firm determines its price behavior based on a prediction about its rivals' reactions to potential price changes. This is one way to inject strategic considerations into the firm's decisions. Paradoxically, the willingness of firms to respond aggressively to price cuts is the very thing that sustains stable prices. Price cuts will not be attempted if they are expected to beget other cuts. Unfortunately, the kinked demand curve model is incomplete. It does not explain why the kink occurs at the price P*. Nor does it justify the price-cutting behavior of rivals. (Price cutting may not be in the best interests of these firms. For instance, a rival may prefer to hold to its price and sacrifice market share rather than cut price and slash profit margins.) A complete model needs to incorporate a richer treatment of strategic behavior.

Price Wars and the Prisoner's Dilemma

Stable prices constitute one oligopoly outcome, but not the only one. In many markets, oligopolists engage in vigorous price competition. To this topic we now turn.

A surprising number of product lines are dominated by two firms, so-called duopolists. Some immediate examples are Pepsi versus Coke, Nike versus Reebok (running shoes), Procter & Gamble versus Kimberly-Clark (disposable diapers), and Disney-MGM versus Universal (movie theme parks). When the competing goods or services are close substitutes, price is a key competitive weapon and usually the most important determinant of relative market shares and profits.

A PRICE WAR As a concrete example, consider a pair of duopolists engaged in price competition. To keep things simple, suppose that each duopolist can produce output at a cost of $4 per unit: AC _ MC _ $4. Furthermore, each firm has only two pricing options: charge a high price of $8 or charge a low price of $6. If both firms set high prices, each can expect to sell 2.5 million units annually. If both set low prices, each firm's sales increase to 3.5 million

BERTRAND PRICE COMPETITION An extreme case of price competition originally was suggested by Joseph Bertrand, a nineteenth-century French economist. Suppose duopolists produce an undifferentiated good at an identical (and constant) marginal cost, say $6 per unit. Each can charge whatever price it wishes, but consumers are very astute and always purchase solely from the firm giving the lower price. In other words, the lower-price firm gains the entire market, and the higher-price firm sells nothing. To analyze this situation, suppose that each firm seeks to determine a price that maximizes its own profit while anticipating the price set by its rival. In other words, as in the previous example of quantity competition, we focus on equilibrium strategies for the firms. (The difference is that here the firms compete via prices, whereas previously they competed via quantities.) What are the firms' equilibrium prices? A little reflection shows that the unique equilibrium is for each firm to set a price equal to marginal cost: P_1 _ P_2 _ $6. This may appear to be a surprising outcome. In equilibrium, P _ AC _ MC so that both firms earn zero economic profit. With the whole market on the line, as few as two firms compete the price down to the perfectly competitive, zero-profit level.

Why isn't there an equilibrium in which firms charge higher prices and earn positive profits? If firms charged different prices, the higher-price firm (currently with zero sales) could profit by slightly undercutting the other firm's price (thereby gaining the entire market). Thus, different prices cannot be in equilibrium. What if the firms were currently charging the same price and splitting the market equally? Now either firm could increase its profit by barely undercutting the price of the other—settling for a slightly smaller profit margin while doubling its market share. In summary, the possibilities for profitable price cutting are exhausted only when the firms already are charging P _ AC _MC and earning zero profits.

The Bertrand model generates the extreme result that price competition, by as few as two firms, can yield a perfectly competitive outcome. It should be emphasized that this result depends on two extreme assumptions— that

(1) All competition is on the basis of price and

(2) The lower-price firm always claims the entire market. We already have seen that quantity competition leads to quite a different outcome. Furthermore, even if price is the most important competitive dimension, market shares are unlikely to be all or nothing.9 In models with some degree of product differentiation, competition leads to price reductions, but equilibrium prices remain above the perfectly competitive level.

When to Cut Price

Pricing has been a focus of attention throughout the first half of this book. Let's step back for a minute and take stock of the factors that dictate changes in pricing strategy, in particular, that call for price cuts.

Changes in Market Demand. The surest rationale for a cut in price is an adverse shift in demand. As we've seen, facing a less favorable demand curve means setting a lower optimal sales target and a lower price. Amid a fall in demand because of a growing recognition of health risks, tanning salons have responded by cutting prices. Seeing buyer demand sapped by the ongoing US recession, Saks Fifth Avenue broke ranks with other upscale retailers by sharply discounting its prices at the start of the 2008 holiday buying season.

Market Skimming. This strategy of price discriminating over time means setting a high price to pioneer adopters (who have relatively inelastic demand), then later lowering the price to attract mass-market users (whose demand is more elastic). Apple's iphone and ipad both saw significant price discounts during their first years on the market.

The Learning Curve. As a firm gains cumulative experience producing a new product, it can expect to reduce its cost per unit by reengineering and improving the production process. Lower unit costs support lower prices. More important, it pays for the firm to cut a product's price at the outset in order to induce a "virtuous circle" of profitability. The initial price cut spurs sales and production levels, speeding the learning process, thereby accelerating cost efficiencies and, in turn, supporting further price reductions—with additional profit accruing to the firm at each stage. Strong learning curve effects have been documented for a range of assembly-line products: from aircraft to laptops to photocopiers.

Strategic Price Cuts. Increased competition from competitors—whether in the form of advertising, quality improvements, or aggressive pricing—can be expected to have an adverse effect on the firm's demand and, therefore, might call for price cuts in response. For instance, Neiman Marcus Group, Gucci, Hermes, and several top fashion houses were compelled (albeit belatedly) to follow Saks's price discounting strategy. Major airlines routinely meet the challenge of a rival introducing additional flights along its routes by offering fare discounts.

Boosting Sales of Related Products. When a firm sells complementary products, cutting the price of one spurs the demand for another, and more importantly, is the path to maximizing the firm's total profit. Gillette is

happy to give away its multiblade razors at minimal cost because the company generates its real profit by selling packs of replacement blades at a price upward of $2 per blade. As long as a consumer is locked into his favorite shaver, the money from blade purchases will keep on coming. Microsoft has long underpriced its Windows operating system because that platform generates significant demand for its applications software such as Microsoft Office. Google generates so much revenue (some $30 billion in 2010) from Internet advertising that it makes sense to tie consumers to Google by giving away free such key online features as e-mail, Google Maps, and its Chrome browser.

The Kindle Once Again. As we saw in Unit 3, since introducing the Kindle in 2007, Amazon has repeatedly cut its price—from $399 to $259 to $189 to $159. Each of the factors listed above has a bearing on this pricing strategy. A skimming strategy certainly makes sense—setting high prices to hard-core egadget aficionados and subsequently lowering prices to enlist the less sophisticated mass market of buyers. So too has there been a steep learning curve, lowering the production cost of the Kindle over time. Moreover, as noted in Unit 6, Amazon has an obvious incentive to lower the Kindle's sale price in order to boost the lucrative tied sales of its ebooks. Additional e-book sales translate directly into greater total profit. Cutting price is also a logical competitive response. Facing increased price competition from Barnes & Noble's Nook reader and the threat of losing users to Apple's multipurpose iPad, Amazon's Kindle price cuts make sense as a profit-maximizing countermove. Of course, a less charitable interpretation suggests that Amazon might be on the verge of becoming ensnared in a destructive price war. Finally, some book publishers have claimed that Amazon CEO Jeff Bezos's real goal is to obliterate the hardcover book market altogether and, simultaneously dominate the emerging e-book market. Such an extreme strategy could mean selling the Kindle at a loss and might be far from optimal. In other words, this last price-cutting explanation owes more to psychologically driven (perhaps irrational) behavior than to profit maximization.

OTHER DIMENSIONS OF COMPETITION

Thus far, our focus has been on quantity and price competition within oligopolies. In this final section, we briefly consider two other forms of competition: strategic commitments and advertising.

Strategic Commitments

A comparison of quantity competition and price competition yields a number of general propositions about the strategic actions and reactions of competing firms. Consider once again the case of symmetric firms competing with respect to quantities. A key part of that example was the way in which one firm's quantity action affected the other's—that is, how the competitor would be expected to react. If one firm (for whatever reason) were to increase

its quantity of output, then the profit-maximizing response of the other would be to decrease its output. (Roughly speaking, the greater is one firm's presence in the market, the less demand there is for the other.) Equation 9.3's reaction function shows this explicitly. We say that the firms' actions are **strategic substitutes** when increasing one firm's

action causes the other firm's optimal reaction to decrease. Thus, the duopolists' quantity decisions are strategic substitutes.

By contrast, price competition works quite differently. If one firm changes its price (up or down), the optimal response for the competing firm is to change its price in the same direction. (One firm's price cut prompts a price cut by its rival. Conversely, if one firm raises its price, the other can afford to raise its price as well.) The earlier example of Bertrand (winner take all) price competition exhibits exactly this behavior. Similar (but less dramatic) price reactions occur when competition is between differentiated products. (Here, a price cut by one firm will attract only a portion of the other firm's customers and so prompts only a modest price reaction.) We say that the firms' actions are **strategic complements** when a change in one firm's action causes the other firm's optimal response to move in the same direction. A comparison of competition between strategic substitutes and strategic complements leads to the following proposition.

In a host of oligopoly models, competition involving prices (strategic complements) results in lower equilibrium profits than competition involving quantities (strategic substitutes).

This result underscores the key difference between firm strategies under price competition and quantity competition. When firms compete along the price dimension, a rival's lower price leads to the firm lowering its own price. In short, competition begets more competition. By contrast, under quantity competition, a rival's increase in output induces a lower quantity of output by the firm itself. In this sense an increase in output deters a competitive response. In general, price competition is more intense than quantity competition (which is self-limiting). The upshot is that equilibrium price setting tends to lead to lower profits for the firms than equilibrium quantity setting.

Advertising

For firms competing in an oligopoly, advertising can be a powerful means of promoting sales. Indeed, firms that sell differentiated goods spend enormous sums on advertising. We begin this section by analyzing a single firm's optimal advertising decision. Later, we consider advertising as a competitive weapon within oligopoly.

OPTIMAL ADVERTISING

Consider a consumer-products firm that must determine not only the price at which to sell one of its goods but also the associated level of advertising expenditure. At a given price, an increase in advertising will raise sales to a greater or lesser extent. One way to picture the firm's decision problem is to write its demand function as Q(P, A). Here the demand function, Q, shows that the quantity of sales depends on price, P, and advertising expenditure, A. The firm's total profit in terms of P and A can be written as

$$\pi = P \cdot Q(P, A) - C[Q(P, A)] - A.$$

Profit is simply revenue minus production cost minus total advertising cost. We see that determining the level of advertising involves a basic trade-off: Raising A increases sales and profits (the net value of the first two terms) but is itself costly (the third term). As always, the optimal

level of advertising is found where marginal profit with respect to A is zero. Taking the derivative of Equation 9.6 and setting this equal to zero, we find

$$M\pi_A = \partial\pi/\partial A = P(\partial Q/\partial A) \quad (dC/dQ)(\partial Q/\partial A) \quad 1 = 0$$

or

$$(P \quad MC)(\partial Q/\partial A) = 1.$$

The left-hand side of this equation is the marginal profit of an extra dollar of advertising, computed as the increase in quantity (Q/ A) times the profit contribution per unit. The right-hand side is the MC of advertising ($1). Optimal advertising spending occurs when its marginal benefit (in terms of profit) equals its marginal cost.

ADVERTISING WITHIN OLIGOPOLY

To consider the impact of advertising in an oligopoly, we must move from a single firm's point of view and ask: What is the effect when a small number of oligopolists simultaneously pursue optimal strategies? To illustrate the possibilities, we briefly consider two polar cases. 1. Product Differentiation. One role of advertising is to underscore real or perceived differences between competing products, that is, to promote product differentiation and brand-name allegiance. Thus, the aim of a firm's advertising is to convince consumers that its product is different and better than competing goods, for example,

"Coke is the real thing," "Only Rolaids spells relief," and "Tropicana Orange Juice tastes like fresh squeezed, not concentrate." From the firm's point of view, the ideal result of such advertising is to create a large segment of loyal consumers—customers who will not defect to a rival product, even if the competitor offers a lower price or enhanced features.

In economic terms, increased product differentiation lessens the substitutability of other goods while reducing the cross-price elasticity of demand. In other words, it tends to blunt competition between oligopolists on such dimensions as price and performance. (For instance, because of heavy advertising, Dole pineapples and Chiquita bananas enjoy much higher price markups than generic fruit.) The individual oligopolistic firm finds it advantageous to differentiate its product. Moreover, the firms' simultaneous advertising expenditures may well result in increased profits for the oligopoly as a whole.[11]

2. Informational Advertising. A second major role of advertising is to provide consumers better information about competing goods. Claims that "We offer the lowest price" (or "best financing" or "50 percent longer battery life" or "better service" or "more convenient locations") clearly fall into this category. Advertising copy frequently provides direct descriptions of products, including photographs. The effect of purely informational advertising is to make consumers more aware of and sensitive to salient differences among competing products. When imperfect information is the norm, some firms might charge higher-than-average prices or deliver lower-than average quality and still maintain modest market shares.

Informational advertising tends to eliminate those possibilities and forces firms to compete more vigorously for informed consumers. The result is lower prices (and/or improved product quality) and lower industry profits.12

Across the spectrum of oligopoly, both reasons for advertising—to differentiate products and to provide information—are important. Both effects provide firms an economic incentive to advertise. (Indeed, only under perfect competition—where products are standardized and consumers already have perfect information—would we expect advertising to be absent.) However, the implications for firms and consumers (whether advertising enhances or blunts competition) tend to work in opposite directions. Not surprisingly, a number of commentators and policy makers have attacked pervasive advertising as anticompetitive. (In novelist F. Scott Fitzgerald's words, "Advertising is a racket. Its contribution to humanity is exactly minus zero.") However, it is mainly an empirical question as to which aspect of advertising—its pro-competitive or anticompetitive effect—tends to be stronger and more important.

There have been numerous research studies concerning the effect of advertising in different industries over different time periods.13 Overall, findings are mixed. Advertising about price has been found to lower average prices for consumer products, such as toys, gasoline, pharmaceuticals, and eyeglasses. (For instance, consumers in states that ban eyeglass advertising pay higher prices than consumers in states that allow it.) There is evidence that advertising (once vigorously fought by state and national bar associations) can lower the price of legal services. In short, in certain markets, advertising plays an important role in providing price information. However, there is also countervailing evidence that advertising and product differentiation can create entry barriers and increase industry concentration and profits.

REVIEW QUESTION

1. What is oligopoly market
2. Describe quantity competition in oligopoly market.
3. Explain about price competition oligopoly market.
4. What is optimal level of advertising in oligopoly market?

FURTHER READINGS

1. Managerial Economics - Arun Kumar, Rachana Sharma
2. Managerial Economics - Thomas J. Webster
3. Managerial Economics - Petersen / Jain
4. Managerial Economics - Yogesh Maheshwari
5. Managerial Economics - E. Narayanan Nadar, S. Vijayan

UNIT-10 GAME THEORY AND COMPETITIVE STRATEGY

CONTENTS

INTRODUCTION

The key presumption of game theory is that each decision maker (or player) acts rationally in pursuing his or her own interest and recognizes that competitors also act rationally.1 Although rational behavior may be directed toward a variety of goals, the usual operational meaning is that all players pursue profit-maximizing strategies and expect competitors to do likewise. (In this sense, the models of quantity and price competition discussed in the preceding Unit are game-theoretic models.)

SIZING UP COMPETITIVE SITUATIONS

A convenient way to begin our discussion is with an overview of the basic game theoretic elements of competitive situations. We begin with elements common to all competitive situations.

1**. Players and Their Actions.** If it is to have a strategic interest, the competitive situation must involve two or more players whose choices of actions affect each other. (It is customary to use player as a catch-all term. Depending on the context, a player may be a private individual, a manager, a firm, a government decision maker, a military leader, a representative of a group or coalition, you name it.) In the example opening this Unit, the players are the managers of three competing airlines. Each must decide what action to take—what number of daily departures to fly along the air route in question. By deliberate intent, this example considers only one kind of action. Generally, an airline's operations on a single air route involve decisions about prices, schedules, plane configurations, in-flight services, advertising, and so on. In broadest terms, an airline strategy would encompass marketing decisions (advertising, the use of computerized reservation systems, and frequent-flier programs), investment decisions (ordering planes, expanding terminals, and choosing hubs), manpower and labor decisions, and merger and acquisition strategies.

2. **Outcomes and Payoffs.** The firm's action, together with actions taken by its rivals, determines the outcome of the competition. In the battle for air passengers, the three airlines' numbers of departures completely determine their market shares (and the number of tickets they sell). Associated with any outcome is a payoff that embodies each competitor's ultimate objective or goal. For a private firm, such as an airline, this payoff usually is measured in terms of monetary profit. In other situations, payoffs take nonmonetary forms. In a war, payoffs might be expressed in terms of territory taken, number of enemy killed, and so on. In the race for the U.S. presidency, payoffs might be counted in Electoral College votes. In short, a payoff summarizes and measures the preferences of a given player.

3. **Underlying "Rules."** Just as important as the players, actions, outcomes, and payoffs are the formal and informal rules that govern the behavior of the competitors. One category of rules includes generally agreed-upon competitive practices, laws, and specific industry regulations. For instance, before 1978, the airline industry operated under strict government regulations. In the current era of deregulation, price and entry constraints have been dropped.

Nonetheless, myriad antitrust rules and regulations prohibit price collusion, unfair practices, and mergers that would increase monopoly power. A second category of "rules" provides a framework to model the competition. They specify whether competitors take actions simultaneously or sequentially. If sequentially, who moves first, second, or last? These rules also describe what each competitor knows about the others' preferences and previous moves at the time it takes action. In the battle for air passengers, airlines set their number of departures independently and without knowing their competitors' decisions. Equally important, competitive situations differ across a number of dimensions.

1. **Number of Competitors.** The number of competitors is one fundamental way to categorize competitive situations. We distinguish between settings with two competitors (so-called two-person games) and those with more than two (n-person or many-person games). In a two-person game, you and your adversary have conflicting interests to a greater or lesser degree. In the preceding Unit, we considered quantity and price competition between duopolists. In Unit 15, we will examine two-party negotiations: between buyer and seller, management and labor, plaintiff and defendant. Frequently, one can analyze multicompetitor settings as if they involved only two parties: the firm in question and all other competitors. This is true in the battle for air passengers. One airline's market share depends on its own number of departures and on the total departures by its competitors (not the particular breakdown). Thus, an airline need only anticipate the average decisions of its competitors to determine its own best response.

2. **Degree of Mutual Interest.** In some situations, the interests of the competitors are strictly opposed; one side's gain is the other side's loss. At the end of a poker game, for example, there is simply an exchange of dollars. Since winnings are balanced by losses, the total net gain of the players together is equal to zero. In the terminology of game theory, this type of competitive situation is called a **zero-sum game**. The zero-sum

game may be thought of as one extreme—that of pure conflict. At the other extreme are situations of pure common interest—situations in which "competitors" win or lose together, and both prefer the same outcome. Real-world examples of either pure cooperation or pure conflict, however, are the exception. In most settings, players exhibit varying degrees of common interest and competition. Because different outcomes can lead to very different (and nonoffsetting) gains and losses for the competitors, these situations are designated **non-zero-sum games**.

3. **Communication and Agreement among Competitors.** In the battle for passengers, the competing airlines make independent decisions. If the battle turns bitter and all airlines set numerous flights, the eventual outcome may well be losses for all carriers. By contrast, if rival airlines were allowed to communicate their intentions and coordinate their operations, one would expect them to agree to mutual flight reductions. (One also would expect cooperation on other competitive dimensions, such as higher prices, less generous frequent-flier programs, and so on.)

4. **Repeated or One-Shot Competition.** Another important distinction is whether the competition is one shot or ongoing—that is, whether the same parties will be involved in similar situations in the future. For instance, competition among airlines is ongoing. Similarly, when management and union representatives negotiate a contract, they recognize that the bargaining will repeat itself three or so years down the road when the new contract expires. By contrast, a buyer and seller negotiating a house sale are unlikely to meet again. In one-shot situations, competitors usually are out for all they can get. In an ongoing competition, they often behave much differently. All they can get now is tempered by the impact on what they might get in the future. If management negotiates too stringent a contract this time, the union may be more militant the next time. As we shall see, if a non cooperative situation is repeated or ongoing, a clear opportunity is provided for tacit communication and understanding to take place over time.

5. **Amount of Information.** The degree of information one competitor has about another is one of the most important factors in a competitive situation. In many industries, secrecy is crucial. Detroit's automakers carefully guard their new designs. At the same time, some firms invest large sums attempting to obtain information about their competitors. Management usually knows who its main rivals are, but it may have only sketchy knowledge of their intentions, views, and ultimate objectives. Normally, the firm has limited information about its competitors' organizations, such as their intentions and costs. This raises the questions: What would management like to know about its competitors? What would management like them to believe about its own intentions?

ANALYZING PAYOFF TABLES

The starting point for a game-theoretic analysis of any competitive situation is a description of the players, their strategies, and their payoffs. Here is a motivating example.

A **dominant strategy** is a best response to any strategy that the other player might pick. Thus, we have shown that scheduling its hit at 8 P.M. is NBC's dominant strategy. By similar reasoning, CBS's dominant

strategy is to lead with its hit. (If NBC schedules its hit at 8 P.M., CBS prefers a 33 million audience to a 28 million audience; if NBC puts its hit at 9 P.M., CBS prefers a 36 million audience to a 30 million audience.) The predicted outcome of the ratings battle is for each network to use its dominant strategy, that is, schedule its hit at 8 P.M.

This results in audience shares of 36 million and 33 million, respectively. As a simple variation on this example, suppose CBS is aware that scheduling its hit against NBC's hit would be suicidal. (Imagine NBC's hit to be the top rated show.) To illustrate, change CBS's top-left entry in Table 10.1 from 33 to 25. How does this change CBS's behavior? Now CBS's best response is to put its hit at 9 P.M. if NBC schedules its hit at 8 P.M. (Of course, CBS's best response is to put its hit at 8 P.M. if NBC schedules its hit at 9 P.M.) In other words, CBS should set its schedule to avoid a showdown of hit shows. CBS no longer has a dominant strategy; rather, its best response depends on what NBC does. Nonetheless, its optimal action is easy to determine. NBC surely will choose to schedule its hit at 8 P.M., because this is its dominant strategy. Anticipating this move, CBS should place its hit at 9 P.M. as a best response. The network outcomes are audiences of 39 million and 28 million viewers, respectively. In this variation on the basic example, CBS's optimal action requires a simple kind of reflexive thinking: putting itself in NBC's shoes. Notice that the predicted outcome has the property that each player's strategy is a best response against the chosen strategy of the other. Thus, neither network could improve its profit by second-guessing the other and moving to a different strategy.

Equilibrium Strategies

What action should a decision maker take to achieve his objectives when competing with or against another individual acting in her own interests? The principal answer supplied by game theory is as follows: In settings where competitors choose actions independently of one another (and so cannot collude), each player should use an equilibrium strategy, one that maximizes each player's expected payoff against the strategy chosen by the other. This

is known as a Nash equilibrium.In both versions of the ratings battle example, the predicted outcome satisfies this definition; that is, it is an equilibrium. The following example illustrates a competitive setting in which neither side has a dominant strategy. Nonetheless, each side has an equilibrium strategy, and that is how each should play.

MARKET-SHARE COMPETITION

Consider two duo polists who compete fiercely for shares of a market that is of constant size. (The market is mature with few growth opportunities.) Each firm can adopt one of three marketing strategies in an attempt to win customers from the other. The payoff table in Table 10.2 depicts the percentage increase in market share of firm 1 (the row player). For instance, if both firms adopt their first strategies, firm 1 loses (and firm 2 gains) two share points. As described, the market share competition is a **zero-sum game**. The competitors' interests are strictly opposed; one side's gain is the other side's loss. This being the case, it is customary to list only the row player's payoffs. The row player seeks to maximize its payoff, while the column player seeks to keep this payoff to

a minimum. By doing so, firm 2 maximizes its own increase in market share.

In the advertising competition, there is a single equilibrium pair of strategies: R2 versus C2. The resulting payoff (two here) is called the equilibrium outcome. To check that this is an equilibrium, consider in turn each firm's options. Against C2, the best firm 1 can do is use R2. Switching to R1 or R3 means suffering a loss of market share. Similarly, the best firm 2 can do against R2 is use C2. If it switches to C1 or C3, it grants firm 1 a greater share increase, implying a greater loss in market share for itself. Thus, the strategies R2 and C2 are profit maximizing against each other and constitute a Nash equilibrium. To check that this is the only equilibrium, let's identify each firm's best response (i.e., its most profitable action) to any action taken by its competitor. Firm 1's best response to C1 is R3, to C2 is R2, and to C3 is R1. Certainly, if firm 1 could anticipate firm 2's action, it would use its best response against it. In Table 10.2, the payoffs from firm 1's best responses to firm 2's possible actions are circled. The circles offer visual proof of the fact that firm 1 has no dominant strategy. (Why? If a strategy were dominant, all the circles would line up along the same row.) The table also identifies firm 2's best responses: Its best response to R1 is C1, to R2 is C2, and to R3 is C3. The resulting payoffs are enclosed in squares. (Firm 2 has no dominant strategy.) The circles and squares make it easy to identify the equilibrium outcome and strategies. A payoff is an equilibrium outcome if and only if it is enclosed by both a circle and a square; that is, it must be a best-response strategy for both players. Thus, we confirm that 2 is the unique equilibrium outcome; R2 versus C2 are the equilibrium strategies that generate this outcome. The best a smart player can expect to get in a zero-sum game against an equally smart player is his or her equilibrium outcome. If either side deviates 406 Unit 10 Game Theory and Competitive Strategy

Competitive Advertising in a Mature Market

In this zero-sum game, the firms' equilibrium strategies are R2 and C2.

		Firm 2		
		C1	C2	C3
Firm 1	R1	[−2]	−1	④
	R2	5	[②]	3
	R3	⑦	−3	[−5]

from its equilibrium play, it reduces its own payoff and increases the competitor's payoff. Indeed, there should be no real uncertainty about how the game will be played. Each side should anticipate equilibrium behavior from the other. The resulting equilibrium outcome is called the value of the game.

		Firm 2		
		C1	C2	C3
Firm 1	R1	43, 57	44, 56	49, 51
	R2	50, 50	47, 53	48, 52
	R3	52, 48	42, 58	40, 60

A REMINDER

It is important to distinguish clearly between a Nash equilibrium that involves dominant strategies and one that does not. Here is the difference: In dominant-strategy equilibrium, each player chooses an action that is a best response against any action the other might take. In a Nash equilibrium, each player takes an action that is a best response to the action the other takes. Both kinds of equilibrium share the essential feature of stability. In equilibrium, there is no second guessing; it is impossible for either side to increase its payoff by unilaterally deviating from its chosen strategy. The concepts differ in one important respect. When a player has a dominant strategy, there is no circumstance in which doing anything else ever makes sense. The player always should use this strategy. Of course, in many, if not most, competitive situations, players will not have available a single strategy that is dominant. However, as in the market-share competition, there still will be a Nash equilibrium. Here each side's action is a best response against the other's. As long as each competitor is smart enough to recognize the Nash equilibrium and expect the other to do likewise, this is how each should play. But what if one player is not so smart? Consider the market-share battle once again. Suppose the manager of firm 2 is convinced that firm 1 plans to use strategy R3. This might not seem to be a very smart move by firm 1. (Perhaps it is lured to R3 by the mistaken hope of a _7 payoff.) But let's say that there is ample evidence that this is how firm 1 will play. (It already has begun launching the R3 advertising campaign.) Then, surely, firm 2 should choose C3, gaining a 5 percent share increase at firm 1's expense. By changing from C2 to C3, firm 2 can profit from firm 1's mistake. The point is this: In a Nash equilibrium (unlike a dominant-strategy equilibrium), there exist some circumstances where it might pay to use a non equilibrium strategy. If one player deviates from equilibrium (by mistake or for any other reason), the other player may be able to improve its payoff by deviating also. Are we recommending non equilibrium play in Table 10.2 for either firm? Certainly not. Equilibrium play is quite transparent and should be grasped readily by both sides. But in a different setting where there is reason to anticipate one player deviating from equilibrium play, the other player may be able to profit from that action by deviating (optimally) as well.

THE PRISONER'S DILEMMA ONCE AGAIN Before concluding this section, we take a brief second look at the paradigm of the prisoner's dilemma (PD) introduced in Unit 9. The top portion of Table 10.3 reproduces the price-war payoffs of Table 9.2. The middle portion of the table portrays a different sort of PD: an arms race between a pair of superpowers. Finally, the bottom portion uses symbolic payoffs to represent the general features of the prisoner's dilemma. Although particular payoffs vary, the strategic implications of the three payoff tables are the same. Assuming non cooperative play (i.e., no possibility of communication or collusion), self-interest dictates the play of dominant strategies. In the price war, a low price is most profitable, regardless of the competitor's price. Similarly, an arms buildup is the dominant strategy in the arms race. (Fortunately, events in the former Soviet Union and the end of the cold war have called a halt to the arms

buildup.) Finally, in the generic prisoner's dilemma, defection is the dominant strategy. Note that the temptation payoff from defecting is greater than the reward payoff from cooperation. In turn, the penalty payoff if both players defect is greater than the sucker payoff if only one player cooperates. In short, the logic of dominant strategies inevitably leads to the inferior penalty payoffs under non cooperative play.

Three Prisoner's Dilemmas

In each case, the play of dominant strategies leads to inferior group outcomes.

(a) A Price War

		Firm 2: High Price	Firm 2: Low Price
Firm 1	High Price	10, 10	5, 12
	Low Price	12, 5	7, 7

(b) An Arms Race

		Superpower 2: Disarm	Superpower 2: Build Arms
Superpower 1	Disarm	10, 10	−50, 20
	Build Arms	20, −50	−20, −20

(c) A Generic PD

		Player 2: Cooperate	Player 2: Defect
Player 1	Cooperate	R, R	S, T
	Defect	T, S	P, P

T = Temptation
R = Reward
P = Penalty
S = Sucker

T > R > P > S

What if the rules of the competition allow communication between players, and what if binding agreements are possible? Under these cooperative ground rules, players should agree to take actions to achieve the mutually beneficial "upper-left" payoffs. Thus, firms would want to agree to charge high prices, and superpowers would strive to negotiate a binding and verifiable arms control treaty. We will say more about the possibilities of reaching such agreements in our later discussion of repeated competition.

A common takeover tactic in the 1980s and early 1990s was the "two-tiered" tender offer. Here is a bare-bones example of how this kind of offer works. Suppose that firm A (the acquiring firm) is seeking to gain control of firm T (the target). Firm T's current (i.e., pre-tender) stock price is $50 per share. Firm A offers a price of $55 per share for 50 percent of firm T's outstanding shares. If 50 percent of shareholders (induced by this price) tender their shares, this percentage will be just enough for firm A to gain control of firm T. In keeping with a two-tier strategy, firm A offers only $35 per share for the remaining 50 percent of shares. Does this two-tiered offer strategy make sense? Will firm A succeed in gaining control? Or will it pay a high $55 price per share but receive only a minority of shares, meaning the takeover will fail? The payoff table below depicts the strategic landscape from the typical shareholder's point of view. The shareholder has two options: to tender her shares or to retain them. The columns show that the shareholder's

payoff depends on her action and on whether or not the acquisition proves to be successful.

		Tender Fails (S < 50%)	Tender Succeeds (S ≥ 50%)
Typical Shareholder	Tender	$55	$45–$55
	Retain	$50	$35

Let's check the payoff entries. The first column entries show that if the tender fails, those who tender shares receive $55, while those who retain theirs see their shares' value remain at the pre-acquisition level of $50. If the tender succeeds, the average price received by those who tender depends on the overall percentage of shares that are offered. If exactly 50 percent of shareholders tender, each tendering shareholder receives $55. At the opposite extreme, what if 100 percent of shareholders tender? Because firm A only buys 50 percent of outstanding shares, each shareholder's offer is prorated, meaning that half of its tendered shares are accepted at $55 and the other half are not accepted. After the acquisition is successfully completed, all remaining shares (i.e., all unaccepted shares) are bought for the lower $35 price. Therefore, the average price received by a typical tendering shareholder is: (.5)(55) _(.5)(35) _ $45. This explains the $45 to $55 payoff range listed in the upper right entry. (Note that if some other percentage of shareholders tendered, say 80 percent, the prorating rule would mean each tendering shareholder would have 50/80 or 5/8 of her shares accepted. Consequently, in a successful tender offer, the average price obtained by any shareholder must lie between $45 and $55 per share. Now that we've anticipated the possible payoffs facing shareholders, the analysis is straightforward. Each shareholder should tender all of her shares, regardless of the percentage of other shareholders who tender. Comparing the entries in the top and bottom rows, we see that tendering is a dominant strategy for every shareholder. (Note that $55 _ $50 and $45–$55 _ $35.) Because every shareholder can be expected to tender, the acquisition easily succeeds and the typical shareholder (after prorating) obtains an average price of $45 for her shares. The extraordinary result is that the acquirer, by structuring a two-tiered offer, pays an average price, $45, which is less than the market value of the target, $50. In other words, target shareholders are getting $5 per share less than what the market deems the firm is worth. Collectively, shareholders are caught in a financial "prisoners' dilemma." They would prefer to hold out for a higher uniform price. But the acquirer has made them an offer that they, individually, can't refuse. Although the two-tiered tender offer has been deemed to be coercive, it has not been found to be illegal. Nonetheless, the majority of U.S. states have enacted rules that effectively restrain the practice, and so with the leveling of the

financial playing field, the two-tiered strategy has all but disappeared over the last 15 years.

COMPETITIVE STRATEGY

Strategic decisions by managers embrace an interesting mixture of competition and cooperation. Firms compete via price wars, patent races, capacity expansion, and entry deterrence. But they also cooperate through joint ventures, the adoption of common standards, and implicit agreements to maintain high prices. The following competitive situations illustrate this blend of competition and cooperation.

A COMMON STANDARD FOR HIGH-DEFINITION DVDS

Holding several times the amount of information, the next generation of digital video disks (DVDs) provides strikingly clear picture quality for movies, video games, and computer graphics. Although the technological hurdles were overcome in 2005 and production began in 2006, a key strategic question remained: Which technology standard and format for DVDs would be adopted in the United States and worldwide?

In one camp, Sony Corporation led a group of companies including Samsung, Matsushita, Philips, Dell Computer, and Hewlett-Packard promoting the so-called Blu-ray format. The opposing side, led by Toshiba Corporation and backed by NEC Corporation and Microsoft, developed the HD format. Each format had its advantages, but each was incompatible with the other. For more than two years, the corporate players formed alliances and pushed their preferred formats. The Blu-ray group enlisted movie studios like Twentieth Century Fox and

Walt Disney. The HD group counted NBC and Universal studios in its camp, and studios such as Warner Brothers and Paramount Pictures pledged to release movies in both formats. Negotiations concerning the standards dispute were overseen by the DVD forum, an industry group made up of some 200 corporate members. However, there was no resolution in sight. The sales of the new DVD players and DVDs lagged; consumers were put off by high prices and, more importantly, by the risk that they might be left with an abandoned technology.

Table 10.4 shows the (hypothetical) payoffs to the two opposing camps associated with the competing standards. Not surprisingly, the Sony group's greatest payoff occurs if all sides adopt the Blu-ray format, whereas the Toshiba group's greatest payoff comes with the HD format. However, coordination is crucial. Both sides receive much lower payoffs if different, incompatible technologies are chosen (the off-diagonal entries).

The payoff table has two equilibria: Both adopt the Blu-ray format (upper-left cell), or both adopt the HD format (lower-right cell). Each is an equilibrium because if one side adopts a given format, the best the other can do is follow suit. (Check this.) Coordination on a common standard is in each side's own best interest. The catch is that the sides have strongly opposed views on which standard it should be. We would expect the outcome to be one of the equilibria— but which one? That is a matter of bargaining and staying power. In general, rational bargainers should agree on a common standard, but such an agreement is far from

guaranteed as evidenced by the actual bitter and protracted dispute. The HD DVD standards dispute was finally resolved in early 2008. The Bluray standard emerged as the winning standard due to a cumulative series of factors. First, Sony installed Blu-ray players in its Play Station 3 game consoles and so attracted video gamers. Second, it gained additional purchase by swaying

The Battle for a Common Technology Standard

The two equilibria have both sides adopting the Blu-ray format or both sides adopting the HD format.

		Toshiba Group: Adopt Blu-ray Format	Toshiba Group: Adopt HD Format
Sony Group	Adopt Blu-ray Format	100, 50	30, 20
	Adopt HD Format	0, 0	60, 90

video distributers, Blockbuster and Netflix, and major retailers such as Best Buy and Wal-Mart to its side. The final tipping point was persuading Warner Bros., the leading video distributer, to release its features exclusively in Bluray. 5 With an overwhelming critical mass of studios, distributers and retailers, the Sony group had effectively claimed the upper-left equilibrium in Table 10.4. Fortunately, mutual advantage is a strong force behind the emergence of common standards. Twenty years ago, there existed a plethora of operating systems in the emerging personal computer market. Today Microsoft Windows is the dominant standard (85 percent market share). More generally, the world has moved toward a number of common standards: metric measurement, left-hand-steering automobiles, and common principles of international law. (Obviously, countries retain different languages, currencies, customs, and laws, even though English, the U.S. dollar, and most recently the euro serve as de facto, partial standards.) Competitive situations such as that embodied in Table 10.4 are ubiquitous. (Standards setting is but one example.) In fact, they commonly are referred to under the label “battle of the sexes.” In that domestic version, husband and wife must decide whether to attend a ball game or the ballet on a given night. Each strongly prefers the other’s company to attending an event alone. The two equilibria have husband and wife making the same choice. But which choice? The wife prefers that they both attend the ball game; the husband prefers the ballet. Based on past experience, we will not hazard a guess as to the outcome of the domestic discussion and negotiations. The general point is that

the battle of the sexes is a model applicable to any bargaining situation. Though it has had its ups and down, your company released its breakthrough product in 2007, a smart phone that combines calling, media playing, and Internet connectivity. A year later came the launch of your online store where users can download tens of thousands of “apps”—applications software enabling the smart phone to do almost anything from playing games to navigating via GPS. You are Steve Jobs, your company is Apple, and the breakthrough product is the iPhone.

Apple’s long time strategic formula has been: “If you build a far better product, they will pay.” Accordingly Apple launched the iPhone at a premium price, and has steadily rolled out improved models (including the iPhone 5) over its first four years. Besides enjoying spectacular sales

and a significant first-mover advantage, Apple wields strict control over the platform—producing the iPhone handset itself, specifying allowable "app" interfaces, approving apps, partnering exclusively with cellular provider AT&T until belatedly adding Verizon as a carrier. In short, Apple tightly controls the iPhone's integrated hardware and software.

Market Entry

Consider once again Unit 1's example of market competition between the two giants of the book business—Barnes & Noble and Borders Group. For two decades, each chain aggressively expanded its number of superstores across the country, often in direct competition with the other. Often the chains were jockeying for the same real estate sites in the same cities. To model the competition between the chains, suppose that both are considering a new superstore in a midsize city. Although the city is currently underserved by the area's bookstores, each chain recognizes that book-buying demand is sufficient to support only one superstore profitably. There is not enough market room for two stores. If both chains erect new superstores and split the market, both will suffer losses. (Each firm's net cash flow will be insufficient to cover the high fixed costs of opening a new store.) Table 10.5 shows the firms' payoffs. If one firm stays out, it earns zero profit. If it enters, its profit
is $4 million or _$4 million depending on whether the other firm enters.
Clearly, neither firm has a dominant strategy. However, it is easy to identify the two off-diagonal outcomes as equilibria. If firm 1 enters, firm 2's best response is to stay out. Thus, entry by firm 1 alone is an equilibrium. By the same reasoning, entry by firm 2 alone is an equilibrium. ("Both firms entering" is not an equilibrium, nor is "both firms staying out." Check this.) Rational competitors should reach one of the equilibria, but it is difficult to say which one. Each firm wishes to be the one that enters the market and gains the profit. One way for a player (say, Borders) to claim its desired equilibrium is to be the first to enter. Here there is a **first-mover advantage**. Given the opportunity to make the first move, Borders should enter and preempt the market. Barnes & Noble's best second move is to stay out. By stealing a march on the opposition—that is, being first to market—a firm obtains its preferred equilibrium. Even if the firms require the same amount of time to launch a superstore,
Borders can claim a first-mover advantage if it can make a credible commitment to enter the market. To be credible, Borders Group's behavior must convince its rival of its entry commitment;6 a mere threat to that effect is not enough. A campaign announcing and promoting the new store would be one way to signal the firm's commitment; another would be entering into a binding real estate lease. Of course, sometimes both firms commit to entry with disastrous results.

Bargaining

One of the most fertile domains for applying game theory is in the realm of bargaining and negotiation. The following example is intended to suggest some of the strategic issues that arise in bargaining settings.

BARGAINING OVER THE TERMS OF A TRANSACTION

Two firms, a buyer and a seller, are in negotiations concerning the sale price of a good. Both sides know that the seller's cost to produce the good is $80,000 and that the buyer's value for the good (the maximum amount the firm can pay) is $120,000. Suppose that, before negotiations begin, each side has formulated its final and best offer, a price beyond which it will not concede in the negotiations. In particular, each is considering one of three possible final offers: $90,000, $100,000, or $110,000.

The firms' offers determine the final price as follows. First, if the firms' price offers are incompatible—that is, the seller insists on a price greater than the buyer is willing to pay—there is no agreement, and each side earns a zero profit. Second, if the players' final offers match, then this is the final price. Third, if the buyer's offer exceeds the seller's demand, the final price is midway between the two offers—as if the players conceded at equal rates toward this final price.

Sequential Competition

In the competitive settings analyzed thus far, players have taken one-shot actions. Of course, many realistic competitive settings involve a series of actions over time. One firm may make a move, its rival a countermove, and so on. In a **sequential game**, players take turns moving. To portray the sequence of moves, we use a **game tree**. As we shall see, when one party makes its current decision, it must look ahead and try to anticipate the actions and reactions of its competitors at their turns in the game tree. To illustrate the method, we start with a compact example.

A multinational firm (MNF) is pondering whether to accept a developing country's (DC) invitation to invest in the development of a copper mine on its soil. Management of MNF is contemplating an agreement in which MNF and DC split the profits from the mine equally. By its estimates, each side's profit is worth about $50 million (in net present value). Both sides are aware that any agreement, being unenforceable, is not really binding. For instance, after MNF has sunk a large investment in the project, DC's leaders could decide to break the agreement and expropriate the mine. Given DC's desperate economic condition, this is a real possibility. In such a case, MNF would suffer a loss of $20 million. The value of the nationalized mine—run less efficiently by DC—would be $80 million. Finally, each side must look to the other to launch the mineral project. MNF sees no other countries in which to invest, and DC has found no other companies capable of launching the mine.

ENTRY DETERRENCE

In the earlier example of market entry, two firms made simultaneous decisions whether or not to enter a market. Let's modify the situation and presume that one firm, the incumbent, already occupies the market and currently holds a monopoly position. A second firm is deciding whether to enter. If entry occurs, the incumbent must decide whether to maintain or cut its current price. The game tree in 10.2a depicts the situation. The new firm has the first move: deciding whether or not to enter. (Because of high fixed costs, entry is a long-term commitment. The new firm

cannot test the waters and then exit.) The incumbent has the next move: maintaining or cutting its price. As the game tree shows, entry is profitable if a high price is maintained but leads to losses if price is cut.

A natural strategy for the incumbent is to threaten to cut price if the new firm enters. If this threat is believed, the new firm will find it in its best interest to stay out of the market. Without a competitor, the incumbent can maintain its price and earn a profit of 12. If the threat works, it will not actually have to be carried out. The beauty of the threat is that the incumbent will have accomplished its goal at no cost. However, the game-tree analysis reveals a significant problem with this strategy. Such a threat lacks credibility. If the new firm were to take the first move and enter the market, the incumbent would not rationally cut price. Once the market has become a duopoly, the incumbent firm's profit-maximizing choice is to maintain price. (A profit of 6 is better than a profit of 4.) In fact, maintaining price is a dominant strategy for the incumbent; high prices are preferred whether or not entry occurs. Thus, the equilibrium is for one firm to enter and the other to maintain price. This example of entry deterrence underscores once again the importance of strategic commitment. If the incumbent could convince the entrant of its commitment to a low price, this would forestall entry. Perhaps one way to accomplish this goal is for the incumbent to cut price before the other firm enters to show its commitment to this low price. If the incumbent can move first and cut its price once and for all, the other firm's best response will be to stay clear of the market. The incumbent certainly would prefer this outcome; its profit is 9, higher than its profit (6) from moving second and accommodating entry. This possibility is depicted in 10.2b's game tree. Be sure to note the reversal in the order of the moves.

Maintaining a lower-than-monopoly price to forestall entry is called **limit pricing**. Cutting price before entry is intended as a signal of the incumbent's price intentions after entry. But is it a credible signal? Again, the real issue is 420 Unit 10 Game Theory and Competitive Strategy commitment. If the incumbent can bind itself to a low-price policy (now and in the future), the new firm will be convinced that entry is a losing proposition. This might be accomplished by making long-term price agreements with customers or by staking the firm's reputation on its low prices. In most cases, however, pricing practices can be undone relatively rapidly and cost lessly. The operative question is, If the entrant were to enter, would the incumbent continue to limit price, or would it revert to a high price that best serves its self interest? If the incumbent is expected to revert, limit pricing loses its credibility and its deterrence effect. Reversion can be depicted by adding a final pricing decision in 10.2b's game tree. Clearly, cutting price in advance does no good if the incumbent is expected to undo the price cut after entry.

BACKWARD INDUCTION

Moving beyond these compact examples, one can construct game trees to model more complicated competitive settings, for instance, those that involve multiple sequential moves by more than two players. As long as the number of moves is finite (so the game cannot go on forever) and all players have perfect information about previous moves, the optimal moves of the players can be found by backward induction, that is, by

solving the game tree from right to left. In other words, to determine a player's optimal action at any point of decision, one must first pin down the optimal plays for all future moves. The resulting sequences of optimal moves constitute the players' equilibrium strategies. Thus, we note an important result in game theory:

Any sequential game with perfect information can be solved backward to obtain a complete solution.

Thinking ahead is the watchword for sequential games. Or, in the words of the philosopher Soren Kierkegaard, "Life can only be understood backwards, but it must be lived forwards."

Repeated Competition

Frequently, firms encounter one another in repeated competition. For instance, duopolists may compete with respect to prices and/or quantities, not just in a single period of time, but repeatedly. Similarly, an incumbent monopolist may encounter a number of would-be entrants over time. How does repetition of this sort affect strategy and behavior?

Repeated competition introduces two important elements into the players' strategic calculations. First, players can think in terms of contingent strategies. For instance, one firm's pricing decision this month could depend on the pricing behavior of its rival during prior months. (The firm might want to punish a rival's price cuts with cuts of its own.) Second, in repeated play, the present isn't the only thing that counts; the future does as well. Accordingly, a player may choose to take certain actions today in order to establish a reputation with its rivals in the future. As we shall see, the use of contingent strategies and the formation of reputations serve to broaden the range of equilibrium behavior.

REPEATED PRICE COMPETITION

As one example of a repeated game, suppose the price competition shown in Table 10.3a is played not once, but repeatedly over time. Thus, when the firms independently set prices in January, they know they will face new price decisions in February and in March and in each succeeding month into the indefinite future. Recall that in one-time play, charging a low price is each firm's dominant strategy. As a result, firms find themselves in a low-profit prisoner's dilemma. But what if the game is played indefinitely? One possibility is for the players to charge low prices every period (that is, simply to repeat the single-stage equilibrium). Charging low prices indefinitely is one equilibrium of repeated competition, albeit a very unattractive one. After all, who wants to be trapped in a prisoner's dilemma forever? Are there other more favorable possibilities? Common sense would suggest that players would strive to coordinate on a cooperative, high-price strategy. The question is how firms can keep this kind of implicit agreement from breaking down. One way is to exploit the power of contingent strategies. Consider the following **punitive** (or grim) strategy:

The firm (1) sets a high price in the first period, (2) sets a high price in every succeeding period, provided the other firm does likewise, and (3) sets low prices forever after, if the other firm ever charges a low price. In short, any defection from the cooperative high-price outcome is penalized by immediate and perpetual defections to low prices. Let's

check that the firms' mutual play of this punitive strategy constitutes an equilibrium in the repeated competition. If each firm adheres to this strategy, each charges a high price in the first and all other periods. Each earns a profit of 10 each period forever. Alternatively, could a firm benefit by unilaterally deviating from the punitive strategy? What if the firm deviated by charging a low price, say in the first period (as good a time as any)? In this period, it increases its profits from 10 to 12. However, this triggers low prices from the other firm forever. Thus, the best it can do is to continue with low prices as well, earning a profit of 7 each period henceforth. Clearly, a one-time 2-unit profit increase is not worth a 3-unit profit reduction into perpetuity.9 Accordingly, the firm's interest is to maintain its reputation for cooperative play throughout the repeated competition. To sum up, the play of punitive strategies, by holding out the threat of retribution, supports a cooperative, high-price equilibrium. The general lesson is that, in infinitely repeated competition, the threat of punishment can be sufficient to enforce a cooperative equilibrium. Indeed, swift but limited penalties may be sufficient to support cooperation. For instance, the strategy "tit-for-tat" is much less drastic than the punitive strategy just described. Under tit-for-tat,

The point of **tit-for-tat** is to deliver a limited punishment for defections from cooperation. If the competitor cuts price one period, the firm cuts its price next period. But if and when the competitor returns to a high price, the firm returns to high prices too. As with the punitive strategies, the mutual play of tit for- tat supports a cooperative high-price equilibrium. With both using tit-for tat, the firms cooperate indefinitely. Neither can gain by a unilateral defection; a one-period gain is not worth triggering an ongoing cycle of defections.

The mutual play of tit-for-tat, or of the punitive strategy, succeeds in supporting a cooperative equilibrium. But these are only two of an endless number of possible contingent strategies. Not surprisingly, there has been considerable research interest in strategies for playing the repeated prisoner's dilemma. An intriguing result of this research is how well tit-for-tat performs in achieving cooperation. Tit-for-tat has four virtues. First, it is nice; it is never the first to defect. Second, it is retaliatory; it immediately punishes an unwarranted defection. Third, it is clear; a competitor can immediately see that it doesn't pay to mess with tit-for-tat. Fourth, it is forgiving; by mimicking the competitor's previous move, it always is ready to return to cooperation. This last feature is the big difference between the punitive strategy (which satisfies the first three features) and tit-for-tat.10

OTHER ASPECTS OF REPUTATION

We have seen that a repeated game allows a player to create and maintain a reputation for cooperation. Reputation can play an analogous role in related contexts. As a simple example, suppose a seller can produce medium-quality goods

or high-quality goods. A typical buyer is willing to pay a premium price for a high-quality item, and the seller could make a greater profit from delivering high quality. The trouble is that the two types of good are indistinguishable at the time of purchase. Only after the buyer has

purchased and used the good is the difference in quality apparent. If only a single, one-time transaction is at stake, we can argue (without needing a payoff table) that the only equilibrium has the seller offering medium-quality goods at a low price. Why? Because a seller's claim for high quality would not be credible. Any buyer who believed the claim and paid a premium price would be exploited by a self-interested seller who delivered medium quality instead.

A FINAL NOTE ON FINITE COMPETITION

We have seen that unlimited repetition can support cooperation in equilibrium. Of course, competition need not go on indefinitely. For instance, one might imagine that there is some probability that the competition will end after any stage. As long as this probability is small enough, the previous analysis, in support of the cooperative equilibrium, continues to hold. However, what happens when the number of periods of competition are limited rather than infinite, that is, when the final period (even one very far in the future) is known? Here the logic of cooperation breaks down. To see this, consider once again the example of price competition played over a fixed number of periods. To find each firm's optimal actions, we work backward. In the last period, each firm's dominant strategy is to cut price, so this is what each does. (No threat of future price cuts can change this because there is no tomorrow.) What about the next-to-last period? With prices sure to be low in the last period, each firm's best strategy is to cut price then as well. In general, if low prices are expected in subsequent periods, each firm's best strategy is to cut prices one period earlier. Whatever the fixed number of periods—3 or 300—this logic carries all the way back to period 1: The only equilibrium is the repeated play of low prices. Thus, we have something of a paradox. When the duration of price competition is limited, super-rational players always will look ahead and see that a price war is coming. Self-interest dictates that it is better to cut price earlier than later. Both sides would prefer high prices, but rational players know that high prices are not stable. Is there a way back to the cooperative, high-price outcome? The answer is yes, if one admits the possibility of near-rational play.

Suppose there is a small chance that one or both sides will play cooperatively because they fail to look ahead to the end of the game. (Perhaps they believe the competition will go on indefinitely.) Injecting this "little bit" of irrationality is a good thing. Now, even a perfectly rational player finds it in his or her self interest to charge a high price and maintain a cooperative equilibrium (at least until near the end of the competition).

REVIEW QUESTION

1. Give a careful explanation of a Nash equilibrium. How is it different from a dominant-strategy equilibrium?
2. Is it ever an advantage to move first in a zero-sum game? When is it an advantage to have the first move in a non–zero-sum game? Provide an example in which it is advantageous to have the second move.

3. Describe the process of analysing payoff tables
4. Discuss about market-share competition
5. What is competitive strategy
6. Describe the process of bargaining over the terms of a transaction

FURTHER READINGS

1. Managerial Economics - Arun Kumar, Rachana Sharma
2. Managerial Economics - Thomas J. Webster
3. Managerial Economics - Petersen / Jain
4. Managerial Economics - Yogesh Maheshwari
5. Managerial Economics - E. Narayanan Nadar, S. Vijayan

UNIT-11 REGULATION, PUBLIC GOODS, AND BENEFIT-COST ANALYSIS

CONTENTS

INTRODUCTION

The role of government in economic decision making encompasses three broad areas: microeconomic, macroeconomic, and distributive. Government's microeconomic role is to provide certain public goods and services, undertake public investments, and regulate operations of private markets. Government's macroeconomic task is to help steer the course of the aggregate economy: reducing the frequency and severity of recessions, promoting economic growth, and maintaining low rates of inflation and unemployment. In its distributive role, government attempts to reduce income inequality; ensure minimum health, education, and living standards; and improve the welfare of the poor. Many government programs serve more than one goal. For instance, a program of increased expenditures may stimulate an economy threatened by recession, redistribute income, and finance spending on particular government budget categories.

This Unit focuses on the microeconomic function of government. In this sphere, the government has two main roles: (1) to regulate private markets by providing basic rules and correcting for market failures that would otherwise result in inefficient production or consumption, and (2) to provide certain desirable public goods and services that are not, or

cannot be, provided via private markets. Part I of this Unit focuses on regulation; Part II applies benefit-cost analysis to evaluate public programs.

I. MARKET FAILURES AND REGULATION

Private markets depend on well-defined property rights. In modern economies, the person on the street sometimes takes property rights for granted. (This is not so in some developing countries, where government politics, judicial failures, and even corruption have hindered the development of private markets.) Private transactions between sellers and buyers start with the seller owning the good and conclude with the buyer obtaining legal possession of the item in exchange for payment. Besides individuals, firms and other organizations have legal rights to undertake transactions, enter into contracts, and create new corporate entities.

Private markets depend on the rule of law created and maintained by government and enforced by the police and the courts. Even today, property rights can change. For instance, tradable pollution permits and rights in genomic material are new types of property rights. Recent court rulings have held that unauthorized downloading music over the Internet violates the property rights of the music's creators. In summary, well-functioning private markets could not exist without the underpinning of government.

In Unit 7, we showed that perfectly competitive markets are efficient, that is, competitive markets provide the right amounts of goods and services at minimum cost to the consumers who value them most highly. This is best thought of as a benchmark. While many markets in the United States approximate the requirements of perfect competition, notable cases of market failure also exist. Market failures usually can be traced to three causes: (1) the presence of monopoly power, (2) the existence of externalities, or (3) the absence of perfect information. In the next three sections, we examine each of these cases in turn.

MARKET FAILURE DUE TO MONOPOLY

Monopolistic markets (pure monopoly, monopolistic competition, or oligopoly) deviate from the efficiency standard. Relative to pure competition, monopoly power elevates prices, increasing the monopolist's profit at the expense of consumer welfare. Consumers lose more than monopolistic producers gain. Thus, total welfare falls.

RENT SEEKING

Because monopoly allows a firm to earn excess profits, companies will invest resources in order to secure a monopoly position. This includes activities directed at the political system (lobbying), the court system (litigation), and the regulatory system (for example, at the Patent Office). Economists call the excess profits that monopolists earn "rents" and call the quest for these rents "rent seeking." Economic theory suggests that firms will compete for rents up to the point where it no longer profits them to do so.

That is, they will compete until most of the excess profits from monopoly have been dissipated through the costs of rent-seeking activity. Rent-seeking activity represents a social loss. (If everyone

stopped doing it, social welfare would increase, even if monopoly remained.) If the monopolist dissipates its excess profits through rent-seeking activity, the total welfare loss of monopoly includes not just the deadweight loss MDE in 9.3 but also the area MBCD. Interestingly, estimates of rent-seeking losses (including resources spent by society to prevent rent-seeking) are typically higher (sometimes much higher) than the estimated deadweight losses due to actual monopolization of markets.2

Government Responses

Antitrust action often is taken to prevent the emergence of monopoly power and restore competition to a monopolistic industry. The U.S. Congress has passed a number of important pieces of antitrust legislation to prevent and attack monopolies. The Sherman Act of 1890 prohibits conspiracies and combinations in restraint of trade, monopolization of any kind, and attempts to monopolize. The Clayton Act of 1914 identifies and prohibits specific types of anticompetitive behavior. The act forbids types of price discrimination aimed at reducing competition in an industry. (Recall that price discrimination occurs when a producer sells the same type of goods to different buyers at different prices.) It also prohibits tying agreements that are used for the purpose of reducing competition. (In a tying agreement, the producer states it will sell a customer a product only if the customer agrees to buy another product from the producer.) The act also prohibits corporations from buying up competitors' shares of stock or having board members in common with competitors if this practice will lessen competition. The Federal Trade Commission Act of 1914 outlaws "unfair methods of competition'' and created the Federal Trade Commission (FTC) to define and enforce this law. In addition, there are a number of other pieces of legislation designed to foster competition. The government can bring suit to enforce the provisions of the various antitrust laws. In addition, both the Sherman Act and the Clayton Act allow private parties who are injured by anticompetitive behavior to bring suit for damages. If successful, the suing party receives three times the value of the actual injury. Suits by either the government or private parties have several aims and results:

1. **Breaking Up Existing Monopolies.** Relying on the Sherman Act, the government may sue to break up a corporation that has attained a monopoly or near monopoly in an industry. In 1911 the government broke up Standard Oil of New Jersey (which controlled over 90 percent of the refining and sales of petroleum products) into 30 independent corporations. In 1982 AT&T, after being sued by the government, agreed to be broken into 23 independent local telephone companies. These operating companies became seven regional phone companies offering local telephone service. The long distance service, Western Electric, and Bell Laboratories were retained in the corporation that kept the name AT&T. Other suits by the government have been less successful. The courts refused to break up U.S. Steel in 1920 and IBM in 1982. In 2001, the Bush administration abandoned attempts to break up Microsoft.

2. **Preventing Monopolistic Practices.** The government seeks to ban practices that firms use (1) to acquire and defend monopoly power and (2) to exploit monopoly power to the detriment of consumers. Such

practices include bundling and tying arrangements, price discrimination, and predatory pricing.

Illegal **predatory pricing** occurs when a large company sets price below cost in order to drive smaller companies out of business. The dominant firm then raises prices once the competitors are driven out. (Companies do not re enter since they know that entry will lead to another round of price cutting.) The problem for courts is to distinguish predatory pricing from virtuous price competition. In 1993 the U.S. Supreme Court cleared Brown and Williamson Tobacco Corporation of predatory pricing charges brought by the Brook

Group, a rival seller of generic cigarettes. The court raised the standard for proving predatory pricing, requiring proof that the accused company deliberately priced at a loss, that this behavior had a reasonable chance of driving rivals out of business, and that the accused would profit as a result. Because of this standard, few predatory cases are brought in the United States and even fewer win. In Europe, where the standard of proof is lower, suits alleging predatory pricing have been more successful. In 2009, the European Commission fined Intel $1.45 billion for offering steep price discounts to customers committing to buying 80 to 100 percent of their needs (allegedly excluding rival suppliers).

3. Preventing Mergers That Reduce Competition. The government also has acted to prevent mergers where the merger would significantly reduce competition. American merger policy was born in opposition to the great wave of mergers and consolidations at the close of the nineteenth century. The original philosophy of the trustbusters was that market dominance and monopoly were bad in and of themselves. Until the 1960s, this remained the prevailing view. In 1962 the government successfully sued to prevent the merger of Brown Shoe and Kinney Shoe, respectively, the fourth and eighth largest manufacturers of shoes at the time. In 1964 the government prevented the merger of the second largest producer of metal containers with the third largest producer of glass containers. And in 1966 the government stopped the merger of two Los Angeles grocery chains that shared just 8 percent of the local market. By the 1970s and 1980s, however, the “Chicago School approach” had assumed dominance in the antitrust arena. According to this philosophy, the forces of free-market competition are far more effective at limiting monopolies than government regulators. Absent prohibitive barriers to entry, a firm’s market power would only be temporary. High profits would attract new entrants attenuating the monopolist’s power. Following this approach, the Reagan and Bush administrations used their antitrust powers sparingly. In the 1990s antitrust thinking accepted new reasons for government action.3 Size was not the first concern. Rather, would the combination have the power to raise prices? For instance, the combination of Staples and Office Depot would have claimed only about 4 percent of the national office supply market. However, the government’s economic analysis predicted that prices would rise by 15 percent or more in markets where the stores formerly competed head to head. Under different circumstances—for instance, if the sixth and seventh largest firms were to combine to compete even handedly with

Likely to Blur under Bush," . the top three firms—mergers could be pro-competitive.

4. **Preventing Collusion.** As we know, firms need not be monopolies to exercise monopoly power. Firms can form cartels and collaborate to reduce output and to fix prices. Such cartels have the same effect on social welfare as do monopolies, and such behavior is illegal. In 1927 the court found that the makers of toilets had acted illegally when they met to fix prices and limit quantities. Even absent an explicit agreement to fix prices, the court may find "conscious parallelism''—that is, a situation in which all producers act in the same way at the same time while being aware that other producers are doing likewise.

MARKET FAILURE DUE TO EXTERNALITIES

An **externality** is a cost or benefit that is caused by one economic agent but borne by another. Pollution is a cost caused by a producer but experienced by others—for example, local residents who suffer deteriorated air quality or immediate neighbors who must endure aircraft noise. Externalities can be negative, as in the case of pollution, or positive. For instance, the pursuit of basic science and research (often government sponsored) generates a host of spinoff benefits to others.

The difficulty posed by externalities is that the party producing the externality has no incentive to consider the external effects on the other, affected parties. The general rule is this: Left to its own devices, the party in question will act so as to produce too much of a negative externality and too little of a positive externality. In short, externalities—either positive or negative—are a potential source of economic inefficiency. To illustrate the externality problem, consider production of a chemical that generates air pollution as a by-product. 11.1 shows the competitive market supply and demand for the chemical. The market equilibrium occurs at the intersection of demand and supply, here at price P_C _ $4 per liter and industry output Q_C _ 10 million liters. In the absence of any externality, this competitive outcome would be efficient.

Suppose, however, that an externality, pollution, is present. To keep things simple, we assume that a known, fixed amount of pollution—say 1 cubic foot of noxious gas—is generated per liter of chemical produced and that each cubic foot causes $1 in harm.

An efficient means of regulation is to tax the producer of a negative externality an amount exactly equal to the associated marginal external cost. In the chemical example, the external cost of pollution is $1 (per extra cubic foot of pollutant), so this is the appropriate tax. In other words, each chemical firm pays a tax, T _ $1, for each cubic foot of pollution it discharges. What is the effect of this tax on the typical chemical producer? By continuing to produce the chemical with pollution as a by-product, the firm incurs an out-of-pocket cost (per additional unit of output) equal to MIC _ T _ MIC _ MEC. Since the tax is set exactly equal to the marginal external cost (MEC), the producer of the externality is made to pay its true social cost. In this way, setting the right tax (T _ MEC) serves to "internalize the externality." With the tax in place, the relevant industry supply curve is MTC (up from MIC, the pretax curve), and the competitive market equilibrium becomes P* _ $5 and Q* _ 8 million liters, precisely the efficient outcome.[7]

Remedying Externalities

The adverse effects of externalities can be ameliorated by a number of means, including (1) government taxes, standards, or permits, or (2) monetary payments between the affected parties established via bargaining or by the courts. We will take up each of these approaches in turn.

We already introduced the argument for imposing taxes and fees on the economic agent causing the externality. Let's take a closer look at the benefits and costs of reducing the externality. 11.2 reconsiders pollution cleanup in its own right, separately from its implications for the output of the chemical industry. As with most activities, the marginal cost rises with increasing levels of cleanup. (The cheapest forms of cleanup are undertaken first.) The marginal benefit of cleanup falls as health gains from cleanup (although positive) exhibit diminishing returns. The optimal amount of cleanup occurs at Q*, where MB _ MC, well short of complete elimination. Beyond this level, the extra benefits are not worth the costs.

The government can promote an output Q* through either pollution fees or quantity standards. The appropriate fee is set at the value of marginal benefit of pollution reduction. Alternatively, the regulator could attain the same result by mandating Q* as the minimum abatement standard. When the regulator has perfect knowledge of the marginal benefit and cost schedules, either regulatory regime can be used to attain the desired result.

In the realistic case of imperfect information, however, externality fees have certain advantages over standards. For example, suppose the regulator is in a good position to estimate the benefits from cleanup but is in the dark about the industry's cost of cleanup. In this case, if the regulator overestimates cleanup costs, the standard will be too lax; if it underestimates these costs, the standard will be too stringent.

Pollution fees, though also subject to error, allow more flexibility. Suppose the regulator mistakenly sets too low a tax; let's say that T _ MB* in 11.2. Since firms clean up only to the point where the marginal cost of doing so equals the tax (MC _ T), the result will be relatively little cleanup. The regulator will see that additional cleanup affords a marginal benefit above marginal cost: MB _ T _ MC. Thus, it can adjust the tax upward until, by trial and error, the resulting level of cleanup satisfies MB* _ T _ MC*, thereby achieving the social optimum.

The advantage of fees over standards is even more pronounced when we recognize the enormity of regulating the myriad sources of pollution. Could a regulatory body, no matter how well informed, be expected to know the marginal benefits and costs associated with each pollution source and set optimal standards? Clearly, such individual standards would be subject to considerable error. In contrast, the value of the tax approach is that all generators of a given externality would be charged the same fee. This uniform fee is set to reflect the estimated externality cost. Whatever their differing costs of abatement, each firm cleans up pollution to the point where its marginal cost equals the tax: $T = MC_1 = MC_2 = \ldots = MC_n$. (Firms for which cleanup is cheap undertake greater pollution abatement.) Marginal costs are equated across all firms,

ensuring that the total amount of pollution is eliminated at least cost. The fee system achieves full efficiency when the tax matches MB*. Regulators have used externality fees in a variety of areas. London, for example, has imposed a congestion fee for driving in the central city. Analysts credit this fee (about $20 per day for driving downtown) with reducing traffic, shortening wait times, and more than doubling average speed in the city. Another regulatory response to an externality, such as pollution, is the introduction of **transferable emissions permits.** The regulator sets the number of permits to allow the discharge of a fixed total quantity of pollution. However, these permits can be bought and sold freely among firms. One would expect a ready market for these permits to emerge. Which firms would end up obtaining and using the permits? Those with the highest cleanup costs. This is exactly the efficient solution the regulator is seeking. A certain amount of pollution is permitted; the rest is cleaned up at least cost.

The trading of pollution permits implies that the required amount of pollution will be cleaned up at least total cost. Nonetheless, the regulator still faces the problem of determining the allowable total amount of pollution (presumably via benefit-cost analysis), and this task is far from easy. Despite these difficulties, the number and scope of emissions trading markets is increasing. In the United States trading permits for sulfur dioxide (the pollutant responsible for acid rain) has been responsible for a 50 percent drop in these emissions. The Chicago Climate Exchange, which operated from 2003 until the end of 2010, allowed corporations to trade greenhouse gas emission credits. In 2008 the Regional Greenhouse Gas Initiative, comprising a dozen U.S. north eastern states, began auctioning off CO_2 permits. The goal is to reduce permits (and thus emissions) by 10 percent by 2018. In 2010, California regulators approved rules to implement the cap-and-trade system established by the state's landmark 2006 Global Warming Solutions Act. Other trading programs have been set up in other parts of the country and overseas. The European Union Emissions Trading System, established in 2005, is the largest carbon emissions trading program in the world.

PRIVATE PAYMENTS

When the affected parties are few in number and property rights are clearly defined, externalities can be resolved efficiently without government intervention.

A classic example is the case of an upstream mill that releases pollutants into a waterway to the detriment of a downstream fishery. Table 11.1 depicts three abatement actions the mill might take and the resulting costs to each party. Of the three options, 50 percent abatement is the efficient solution because this minimizes the total cost incurred by the parties. How might this result actually come to pass? The Coase theorem (developed by Ronald Coase) provides a simple answer: Bargaining between the affected parties will result in an efficient outcome, regardless of the property-rights assignment.[10] To illustrate, suppose the fishery has the right to clean water. Absent any other agreement, it could demand 100 percent cleanup. However, a quick check of Table 11.1 demonstrates the mutual advantage of an agreement at 50 percent cleanup. The mill saves $70,000 in cleanup costs, while the loss to the

fishery is only $30,000. Thus, a payment of, say, $50,000 from mill to fishery in exchange for the right to 50 percent discharge would be mutually beneficial. Suppose, instead, that the mill has the right to pollute (i.e., to elect 0 percent cleanup). Now the fishery must pay the mill to reduce its pollution. Nonetheless, the efficient agreement remains at 50 percent cleanup. A payment by the fishery of $60,000 (or, more generally, any payment between $50,000 and $70,000) would be mutually beneficial. No matter where they start, the parties always have an economic incentive to negotiate their way to an efficient outcome, because this outcome affords the greatest joint gain. Another solution to the problem is to give the party harmed by the externality the right to sue for damages. If an externality is produced, the injured party brings the case to court and will be awarded monetary damages (from the defendant) equal to the economic cost it suffers. This system of private damages is exactly analogous to an externality tax. The initiator of the externality is made to pay the full external cost of his or her actions. The difference is that the payment is private; it goes to the injured party, not the government. As an illustration, suppose the fishery holds the right to clean water and can sue for full damages. The mill has three options: 100 percent cleanup at a cost of $120,000 and no damages paid; 50 percent cleanup at a cost of $50,000 and damages of $30,000 (the harm done to the fishery); or 0 percent cleanup and damages of $100,000. Clearly the mill's cost-minimizing action is 50 percent cleanup. This is precisely the efficient outcome. The world is faced with an environmental problem of unprecedented complexity. Across the globe, countries contribute to global warming through the emission of greenhouse gases (GHG)—primarily carbon dioxide (CO_2), but also methane and nitrous oxides. Sources of these gases include fossil-fuel energy use, industrial and agricultural processes, and forest burning. Surface temperatures have risen almost 1.5 degrees Fahrenheit over the past century, and the rate of increase has increased in recent years. One potential result of warming is a rise in sea levels, implying significant coastline regression across the globe. A second result is regional climatic changes (caused chiefly by alterations in global wind patterns and ocean currents)—less rainfall in the American Midwest and central Canada, more frequent typhoons in the Indian peninsula, possible alteration of the Gulf Stream, reduced water levels in the world's major rivers. A third consequence (largely unknown) is the effect of a CO_2-rich atmosphere on agricultural yields (crop growth, plant diseases, and so on).

Global Warming

A simple fact explains why international cooperation is necessary to address the problem of global warming. GHGs entering the atmosphere from any particular point source are distributed equally around the globe within 12 months. The degree of global warming depends on the total amount of GHGs, regardless of their source. The atmosphere, like many other environmental resources (the open seas, fishing stocks, endangered species), belongs to no countries and all countries. Thus, global warming represents the ultimate externality.

Countries can reduce emissions by a variety of means: reining in heavy industry (at the cost of reducing the rate of economic growth), using

cleaner energy sources (including nuclear power), adopting more fuel-efficient technologies, turning toward greater conservation measures, and replanting forests. However, all of these measures are costly. It is in no single country's interest to institute unilateral reductions in GHGs. Yet, all countries potentially could benefit if multilateral reductions were undertaken.

In principle, the solution to the global warming problem is the same as for any externality. The externality (in this case, total GHG emissions) should be reduced up to the point where the marginal benefit (in terms of a cooler Earth) from any additional reduction just matches the marginal cost (the cost of reducing emissions, including possibly reduced economic growth). Indeed, starting with the 1992 Environmental Summit in Rio de Janeiro, the nations of the world have explored targets and timetables for global emissions reductions. An alternative means of achieving efficient reductions is the implementation of a global "carbon" tax, whereby fossil fuels, automobile emissions, and the like are taxed according to the amount of CO_2 they contribute to the atmosphere. However, two aspects make the global warming problem particularly difficult. The first is the uncertainty about the magnitudes of benefits and costs. Some policy makers call for significant GHG cuts (25 to 40 percent) by 2050, emphasizing the large benefits of reducing global warming and manageable costs. Other experts call for modest reductions, pointing out that the cost of reducing emissions beyond 15 to 25 percent increases exponentially. As yet, there is no consensus on the optimal amount of GHG reductions. The second problem is distributional. The wealthy, industrial countries tend to place the highest value on environmental preservation. (After all, environmental protection is a normal good; as income increases, more of it is desired.) These countries also produce the lion's share of emissions. However, many of the opportunities for low-cost emission reductions reside in the developing world. Thus, there is a mismatch: the developed world lacks the opportunities for low-cost reductions, while the developing world lacks the financial resources to pay for reductions. Thus, payments (or other forms of aid) from industrial nations to developing ones would seem to be a prerequisite for a worldwide reduction plan.

Promoting Positive Externalities

A positive externality occurs when a particular activity has beneficial side effects on parties other than those producing the activity. For instance, efforts to improve literacy and education levels in a particular segment of the population benefit not only the individuals themselves but also society as a whole. By limiting the onset and spread of disease, vaccination programs protect the general population, including those who are not vaccinated.

Left to their own devices, economic agents in unregulated markets tend to undertake too few activities that generate positive externalities. (This is simply the converse of the previous proposition that agents generate too much negative externalities.) The appropriate government intervention is either to mandate or subsidize greater levels of these beneficial activities. In the United States, education is publicly provided and is mandatory through certain grade levels. Similarly, vaccinations

against common diseases can be obtained for free and are mandatory. The following example illustrates the use of subsidies to promote beneficial activities.

PROMOTING RESEARCH

In the United States, private universities and firms undertake the vast majority of basic research leading to new scientific and technological knowledge. As a concrete example, consider a firm engaged in basic research that is contemplating embarking on an R&D program to produce a superior flame-retardant fabric. The firm estimates the expected gross profit of the program (in present-value terms) to be $12 million. It also recognizes that the program will generate external benefits to society as a whole (to consumers and other firms who develop copycat fabrics). These external benefits come to an estimated $6 million. Finally, the firm's total cost of undertaking the R&D program is $15 million.

As far as the firm is concerned, the program's net profit is 12 _ 15 __$3 million. Thus, the firm will choose not to undertake the program. Taking account of total benefits, however, the program should be undertaken. (In total, net benefits come to 12 _ 6 _ 15 _ $3 million.) Clearly, the profit motive alone is not enough to induce the firm to go ahead. What incentive is needed? Simply stated, the government should offer the firm a "carrot,'' that is, an R&D subsidy. What kind and magnitude of subsidy? The answer is straightforward. The crux of the externality problem is that the firm faces paying the entire cost of the R&D program but reaps only two-thirds of the total benefit ($12 million of the $18 million total). Accordingly, the remedy is a "one-third'' subsidy. For every $1.00 of the firm's R&D expenditures, the government reimburses or pays for $.33. With the subsidy, the firm's net R&D cost becomes (2/3)(15) _ $10 million. Therefore, its net profit becomes 12 _ 10 _ $2 million, and the firm elects to undertake the program. The general rule (of which this example is a specific case) is this: To induce efficient behavior, the subsidy should be set equal to the ratio of external benefit to total benefit.

THE PATENT SYSTEM

In the United States, patent law grants the holder exclusive rights to an invention for 20 years. An invention must take the form of a product or process. Intangible knowledge (say, a mathematical theorem) is not patentable. Moreover, the invention must contain a minimum degree of novelty. A mere improvement does not constitute a patentable invention. At the time the patent is granted, the invention becomes public knowledge. What is the economic rationale for patent laws? Their most important role

is to provide incentives for firms (and individuals) to pursue inventions and innovations. Absent patent protection, why should an inventor work to develop an invention, or why should a firm incur the costs to bring it to market? If one did, another firm could duplicate any successful invention and so profit at the expense of the inventor. Without patent protection, a firm that creates an invention would be able to claim only a small portion of the profit generated by the invention. Patent protection encourages the process of invention by allowing the inventor to capture a

greater portion of the benefits created. Patent laws represent a trade-off. On one hand, they provide strong incentives for research and invention in the first place. On the other hand, the patent grants the successful inventor a monopoly over the sale of knowledge embodied in the invention. Like any monopolist, the inventor will set a high price to maximize his or her profit. Because some would-be customers will be unwilling to pay this monopoly price, the knowledge will not be as widely used as it might be.12 As in any monopoly, there is a deadweight loss due to under provision. To sum up, patent protection represents a trade-off between encouraging invention before the fact and disseminating knowledge after the fact. As a practical matter, patent laws do not provide complete protection against imitation. Copycat firms frequently succeed in making just enough changes in the process or product to avoid patent infringement. Nonetheless, patents make imitation more difficult and more costly. Thus, by bequeathing the firm a partial monopoly, patent laws provide a positive profit incentive for invention.

COPYRIGHT

Copyright law provides protection for expressive works, such as music, drama, literature, film, and even software. The Copyright Act of 1790 protected material for 14 years, renewable for another 14 years while the author was still alive. By 1998, this protection had been raised to the life of the author plus 70 years.

Media attention continues to focus on the contours of copyright law, especially in the area of music where technological advances (video recorders, DVD recorders, and the like) have made copying and downloading easy and inexpensive. In the late 1970s, Universal Pictures and Disney sued Sony and other makers of video recorders to prohibit the sale of the recorders, alleging that such recording violated copyright law. The federal district court ruled in Sony's favor and allowed the production of these devices. On the other hand, in 2001, the Court of Appeals (Ninth Circuit) upheld an injunction that effectively shut down Napster, a popular music and file exchange service. In that case, the court objected to Napster making copyrighted songs available from its main server. (So far, courts have refused to prevent the distribution of software that allows direct file sharing among users without the benefit of a central server.) Although the music industry has recently begun suing individual violators, illegal downloads continue to dominate the market. To further fight the downloading problem, the industry has aggressively promoted paid download services such as iTunes, MusicMatch, Rhapsody, and even a new paid Napster. Efficient regulation depends on a careful consideration of benefits and costs.

Regulatory reforms in the 1980s and 1990s have made slow but steady progress in this direction.13 Initially, benefits were not explicitly traded off against costs. For example, the 1970 Clean Air Act specifically excludes a consideration of costs in setting air-quality standards, and the Food and Drug Administration is not obligated to use benefit-cost tests in ascertaining product safety. However, over time regulatory agencies have increasingly turned to a comparison of benefits and costs.

One important area of reform is deregulation.[14] Critics have pointed out that, by intention or not, regulation frequently reduces true competition: Regulated rates can hold prices up as well as down. Critics argue that regulators are often "captured'' by the firms they are supposed to regulate and that government intervention has spread into many areas that are a far cry from natural monopolies: trucking, airlines, and banking, for example. Regulations that limit market entry and fix prices frequently do more harm than good in markets where competition otherwise would be viable.

Beginning in the late 1970s, policy makers have increasingly adopted regulatory reforms calling for deregulation. Deregulation focused on a wide variety of industries, including airlines, banking, brokerage firms, cable television, natural gas, railroads, trucking, and telecommunications. Did the predicted benefits of deregulation come to pass? On balance, the answer is yes. For instance, in the railroad and trucking industries, firms have engaged in vigorous price competition and have become more efficient once free of restrictive regulations. Competition also has been vigorous in the areas of banking and brokerage services. Perhaps most successful has been the case of airline deregulation. Deregulation has produced entry by no-frills airlines, greater competition along high-traffic routes, lower average fares, greater variety and frequency of service, and increased airline efficiency (stemming from hub-and-spoke operations and reduced labor costs) with some reduction in service quality. Overall, consumers have benefited significantly from the 25 years of airline deregulation. Nevertheless, deregulation is not without problems. In 2007 and 2008, the subprime mortgage crisis, partially the result of deregulation, roiled financial markets. The result has been thousands of painful foreclosures, a tightened credit market, lower economic growth, and increased unemployment. Thus government policy makers are revisiting the regulation–deregulation debate. Recently, Congress has enacted broad legislation to more tightly regulate financial markets and institutions.

MARKET FAILURE DUE TO IMPERFECT INFORMATION

In our previous discussion of market efficiency, we took for granted that the consumer is the best judge of the value he or she will enjoy from the purchase of a good or service; that is, the buyer fully understands the benefits and costs associated with any transaction undertaken. This is a good working presumption for many, if not most, transactions. However, some economic transactions involve significant uncertainties as to product quality, reliability, or safety. In these cases, consumers may not have sufficient information to make efficient choices. There are numerous cases of market inefficiencies due to imperfect information, ranging from the routine to the dramatic. As a simple example, consider two lines of household batteries marketed by competing firms. The first firm's battery is a best seller; it is cheaper to produce and thus carries a lower price (10 percent lower) than the competition. According to objective tests, however, the second firm's battery lasts 18 percent longer on average. If consumers possessed perfect information about the batteries, the second battery brand could well be the better seller because

it delivers more power per penny. However, only a minority of consumers (perhaps diligent readers of Consumer Reports) are knowledgeable about the lives of different brands of batteries. Most consumers decide mainly on initial purchase price. Thus, in the presence of imperfect information, a free competitive market will have no way to rid itself of the less-efficient product.

Much more serious examples of market failures occur in the realm of product safety. For instance, consider a hypothetical (or perhaps not so hypothetical) children's toy, a miniature missile launcher. Let's say the toy already is popular in Europe, where it was first marketed. The European experience suggests that the rocket has produced a large number of serious injuries and even near fatalities. In such a case, the prudent regulatory response may be to ban the product in the United States altogether.

When consumers possess imperfect information or misinformation, market outcomes typically will fail the efficiency test. Consequently, there is a potential role for the government. Government regulators, with superior information, may be able to mandate better outcomes than would occur in an unregulated market. Under this rationale, the government bans some drugs, taxes alcohol and cigarettes, mandates compulsory education up to a certain grade, and prohibits the sale of unsafe products. Government can also act by requiring producers to provide certain types of information, such as nutritional labeling on foods or warning labels on cigarettes and wine. Recently, regulations have required restaurants in Los Angeles and New York City to post letter grades reflecting their inspection results in their front windows. The impact has been strongly positive. Restaurant operators have redoubled their efforts to raise cleanliness and food-safety standards.

At the same time, government regulation is not always an ideal remedy. Frequently, the choice is between imperfect markets and imperfect regulation or, sometimes, between market failure and regulatory failure. For instance, the automobile is probably the single most regulated product today. Regulations govern general performance, reliability, safety, fuel economy, and emissions. The majority of these regulations represent improvements over what would be offered in an unregulated market. But almost all these regulations are costly, and not all constitute unambiguous improvements. Later in this Unit, we will pay special attention to how the discipline of benefit-cost analysis can be used to evaluate when and how to regulate for maximum advantage

Business Behavior: Assessing Risks

The modern world is full of more and more things to worry about: global warming, earthquakes, asbestos in buildings, hazardous chemicals, toxins in fish, just to name a few. In making informed decisions—whether to choose air bags on a new car, use lawn pesticides, or go skiing—consumers must grapple with risk assessments all the time. Psychologists have questioned ordinary people to see how accurately they can gauge risks. In a classic study, the psychologist Paul Slovik asked 15 national experts and 40 members of the League of Women's Voters to rank the everyday risks listed in Table 11.2. You can use the alphabetically ordered list of activities in the table to test your own "risk aptitude."

Before turning to Table 11.3, rank the items in Table 11.2 from 1 to 30 in descending order of risk. (In constructing your ranking, consider the total risk to society of the activity or technology.)[15] Table 11.3 lists the activities and compares the rankings of experts and ordinary people (members of the league). Scanning the list, one observes some general level of agreement between the two rankings. Certain items (handguns,motorcycles, smoking) are ranked as high risk and others as low risk (antibiotics, home appliances, power mowers) by both groups. More interesting, however, are the gaps between people's risk perceptions and experts' judgments. Psychologists have found that several factors affect the average person's risk perception. Risks that loom largest in people's perceptions are those that are most visible, imposed (rather than voluntary or under one's own control), man-made (rather than natural), and potentially catastrophic (rather than mundane). For these reasons, the average person tends to overstate the risks from nuclear power, hunting, mountain climbing, skiing, private aviation, and police work. The same person tends to understate the risks from swimming, X-rays, contraceptives, and food preservatives. Compare your own rankings to those in Table 11.3. Were your assessments closer to those of the average person or to those of the experts?

BENEFIT-COST ANALYSIS AND PUBLIC GOODS PROVISION

Benefit-cost analysis is a method of evaluating public projects and programs. [16] It is used in planning budgets, building dams and airports, controlling disease, planning for safety, spending for education and research, and evaluating the costs and benefits of regulation. In short, almost any government program is fair game for the application of the benefit-cost approach. We begin by discussing the economic rationale for the government's provision of certain kinds of public goods. We then go on to outline the basics of benefit-cost analysis.

PUBLIC GOODS

Why are some goods and services provided by government rather than by private markets? What features characterize public goods? A **pure public good** is one that is nonrival and nonexclusive. Roughly speaking, it can be said that "if anyone enjoys the public good, everyone enjoys it.'' We can think of a pure public good as the extreme case of an externality: All benefits are external. The prototypical example of a pure public good is national defense. Defense is nonrival; that is, all citizens within the protected area enjoy the benefits of defense. (One state's enjoyment of national defense does not subtract from another state's enjoyment.) Furthermore, national defense is nonexclusive: It is impossible (or certainly impractical) to single out and exclude a particular town or region from the national defense network. A considerable range of other goods, from local police protection to municipal mosquito abatement, share these two properties of pure public goods. Whether or not it is exclusive, a nonrival public good has the feature that increased benefits can be provided to additional people at zero (or negligible) marginal cost. An uncongested highway or bridge has this property. The marginal cost of additional users is zero or nearly

so. Even though exclusion is feasible, it should not be employed. As we shall see, the greatest collective benefit occurs when the highway is toll-free. At a price of zero, no one is excluded and usage is maximized at no additional cost.

Public Goods and Efficiency

Under the basic benefit-cost rule, the government should undertake a project or program if and only if its total benefits, summed over all its users, exceed its total costs. Thus, a stretch of highway should be built if the collective benefits to users (discounted over the course of its life) exceed its total costs—the cost of land taken, highway construction, and annual maintenance. We can refine the question of whether or not to build a highway: What is the optimal size highway to build? Here, we take highway "size'' to mean length in miles. A longer span of multilane, high-speed roadway delivers faster and more numerous trips to more destinations but at an additional construction cost. Consider the planning problem depicted in 11.3. The horizontal axis lists highways of different lengths (in miles) that might be built. The MC curve shows the marginal cost (in millions of dollars) of constructing additional miles of highway. The also presents demand curves for highway trips for two distinct groups: commercial users (business trucks, vans, and the like) and non commercial users ("ordinary'' drivers). Each demand curve measures the marginal benefit for the group from the greater number of trips (and greater convenience) afforded by extra miles of highway.

Identifying a highway of optimal size turns on a comparison of marginal benefit and marginal cost. The key point to recognize is that the total marginal benefit to the groups together is found by taking the vertical sum of the separate marginal benefit (demand) curves. For instance, according to 11.3, a 10-mile-long highway delivers a marginal benefit of $1.75 million per mile to commercial vehicles and $1 million per mile to ordinary drivers. Since these trips are nonrival (i.e., the highway has more than enough capacity for both groups), the total marginal benefit is $2.75 million. More generally, the uppermost "demand" curve shows the sum of the groups' marginal benefits by size of highway. We can now determine the optimal size of the public project in the usual way. In the , a 17.5-mile highway generates the maximum social net benefit. At this size, total marginal benefit equals marginal cost. Two observations are in order. First, there is the problem of financing the project. As pointed out earlier, to optimize usage (and therefore benefits), the highway should be toll-free.[17] Consequently, highway costs must be paid through taxes or government borrowing.

Second, it is difficult to estimate accurately marginal benefits. A sample of commercial and non commercial users can be canvassed concerning their potential usage and value. However, these results are subject to error. The sample may be unrepresentative, and potential users may deliberately misrepresent their values. Intensive users, eager for the highway to be built (and knowing it will be collectively financed), have an incentive to overstate their values. Infrequent users have the incentive to understate their values—to report zero or even negative values—to block spending on the highway. To the extent that marginal benefits (and

marginal costs) are in error, so, too, will be the provision of the public good.

Not surprisingly, spending decisions on public goods frequently are determined as much by politics as by benefit-cost analyses. For instance, the highway decision could be voted on directly by state representatives. The virtue of voting is that it is broadly representative of constituents' preferences. However, many well-known, unavoidable difficulties are encountered with systems of voting. Voting often leads to inconsistent results and, in some circumstances, is subject to undue influence, or even manipulation, by interested parties—all within its ground rules. In addition, a voter's ballot, yea or nay, cannot reflect the magnitude of the individual's true benefit or cost from the project. Thus, a project may receive majority approval even though the dollar gains of the majority fall well short of the total cost incurred by the minority. Conversely, an economically worthwhile project with benefits diffused over a vast, nonvoting constituency may well be blocked by a special interest group that gets out its vote.

THE BASICS OF BENEFIT-COST ANALYSIS

It is best to think of benefit-cost analysis in three steps. For a given course of action, the method (1) identifies all impacts (pro and con) on all affected members of society; (2) values these various benefits and costs in dollar terms; and (3) recommends undertaking the program if and only if doing so produces a positive total net benefit to society—that is, if and only if total benefits exceed total costs.

Applying the Net Benefit Rule

According to the third step in benefit-cost analysis, the decision of whether to undertake a given program hinges on the project's net benefit. The program should be undertaken if and only if Net benefit=Total benefit- Total cost, that is, only if total benefit exceeds total cost. (As we shall see, if benefits and costs occur over time, we must calculate the present discounted value of each using an appropriate rate of interest.)

We can extend this basic rule to the case of several mutually exclusive public programs. For instance, suppose the Department of the Interior is considering building a dam along a major river in the Pacific Northwest. The dam can be built in one of two locations, according to one of three designs. Thus, there are six possible dam plans: seven alternatives, including the option of not building. Among these mutually exclusive alternatives, the one with the maximum net benefit should be selected. (If all dam plans imply negative net benefits, not building the dam delivers the highest net benefit, namely zero.)

A second variation on the basic rule is applicable to public investment decisions involving resource constraints. Suppose that if the dam is built, it will generate 1.5 million acre-feet of water per year. This water can be employed in a number of competing uses, including allocation to city residents, local industry, or farmers, among other segments. From a benefit-cost point of view, the water should be allocated in a way that maximizes total net benefit. A simple rule for allocating the limited supply of water is to compute the net benefit per acre-foot of water in each use. For instance, suppose the city's net benefit comes to $100/acre-foot, industry's to $120/acre-foot, and farmers' to $60/acre-foot. Then

industry's demand should be satisfied first, followed by the city's demand, and finally the farmers' demand.

Efficiency versus Equity

The third step underscores a fundamental tenet of benefit-cost analysis: that only total benefits and costs matter, not their distribution. Thus, a program should be undertaken if it is beneficial in the aggregate, that is, if its total dollar benefits exceed total costs. But what if these benefits and costs are unequally distributed across the affected population? After all, for almost any public program there are gainers and losers. (Indeed, any citizen who obtains no benefit from the program is implicitly harmed. He or she pays part of the program's cost either directly via higher taxes or indirectly via reduced spending on programs the person would value.) Shouldn't decisions concerning public programs reflect distributional or equity considerations?

Benefit-cost analysis justifies its focus on efficiency rather than equity on a number of grounds. The first and strongest ground is that the goals of efficiency and equity need not be in conflict, provided appropriate compensation is paid among the affected parties. Consider a public program that generates different benefits and costs to two distinct groups, A and B. Group A receives a benefit of $5 million; group B suffers a loss of $3 million. The immediate impact of the project is clearly unequal. Nonetheless, if the gainers pay the losers, both groups can profit from the program. The requisite payment must exceed $3 million but not exceed $5 million.

The potential for mutually beneficial compensation exists as long as the program's total net benefit is positive. There are myriad instances in which compensation is paid. For instance, the extension of a desperately needed highway (which would generate significant regional benefits) inevitably means taking land and private homes by eminent domain. Compensation for these losses is accomplished by paying the owners fair market value for the properties. Yet compensation is the exception rather than the rule. In the vast majority of public programs, winners do not compensate losers at all.

The second argument for ignoring equity relies on a form of division of labor. Distribution is best addressed via the progressive tax system and through transfer programs that direct resources to low-income and other targeted groups. According to this argument, it is much more efficient to use the tax and transfer system directly than to pursue distributional goals via specific public investments. Blocking the aforementioned project on equity grounds has a net cost: forgoing a $5 million gain while saving only $3 million in cost. Redistribution via taxes and transfers conserves dollars; there is no net loss. But, of course, how well the tax system addresses distribution problems is open to debate. A third argument in the efficiency-equity debate focuses on the aggregate impact of applying the benefit-cost rule over many projects. The contention is that by following this rule—that is, undertaking only net beneficial projects— long-run total benefits are maximized and project-specific inequities will tend to even out. Clearly, this last contention is an empirical issue. We make one final observation about the efficiency-equity debate. Although it is not common practice, benefit-cost analysis

nonetheless is amenable to the introduction of distributional issues. As step 1 indicates, the benefit-cost method identifies, untangles, and disaggregates the various benefits and costs of all affected groups. This, in itself, is an essential part of making distributional judgments. In standard benefit-cost analysis, when costs and benefits are added, all groups' benefits or costs carry equal dollar weight. One could, however, employ unequal weights to account for distributional concerns. For instance, if group B in the preceding example consists of low-income residents, their dollars might be accorded twice the weight of group A's dollars. With these weights, the benefit- cost analysis now becomes $5 _ (2)($3) __$1 million. Thus, the program would not be implemented because of its effect on distribution. A similar distributional analysis would support investing in a program (even if its net benefit is negative) if its benefits accrue to the neediest in society and its costs fall on the most affluent.

EVALUATING A PUBLIC PROJECT

In this section, we apply benefit-cost analysis to a public investment decision: building a bridge. The decision is not simply whether to invest in the bridge or save one's money. Instead, there are other questions: Is the public investment better than the alternative of regulating the private transport market? Would private investment and control of the bridge be a still better alternative?

Public Investment in a Bridge

A task force of state and city planners is considering the construction of a harbour bridge to connect downtown and a northern peninsula. Currently, residents of the peninsula commute to the city via ferry (and a smaller number commute by car, taking a slow, "great circle" route). Preliminary studies have shown there is considerable demand for the bridge. The question is whether the benefit to these commuters is worth the cost.

The planners have the following information. The ferry currently provides an estimated 5 million commuting trips annually at a price of $2 per trip; since the ferry's average cost per trip is $1, its profit per trip is $1. The immediate construction cost of the bridge is $85 million. With proper maintenance, the bridge will last indefinitely. Annual operating and maintenance costs are estimated at $5 million. Plans are for the bridge to be toll-free. This will price the ferry out of business. The planners estimate that the bridge will furnish 10 million commuting trips per year. The discount rate (in real terms) appropriate for the project is 4 percent. Based on this information, how can the planners construct a benefit-cost analysis to guide its investment decision?

PUBLIC PRICING

Here's a point that should not be overlooked: The decision to build the bridge crucially depends on charging the "right" toll. In the present example, no toll is charged. The right price is zero because there is a negligible cost (no wear and tear or congestion) associated with additional cars crossing the bridge. Thus, a zero price ensures maximum usage. Setting any positive price would exclude some commuters and reduce net benefit. But what if there were significant costs associated

with additional use of a public good? The general principle behind optimal pricing is simple: The optimal price should just equal the marginal cost associated with extra usage. For instance, because large tractor-trailer trucks cause significant road damage to highways, they should pay a commensurate toll. In general, user fees should be set at a level that just covers the marginal cost of the service being delivered.

VALUING BENEFITS AND COSTS

The main issues with respect to valuing benefits and costs concern the role of market prices and ways of valuing nonmarket items.

Market Values

In most cases, market prices provide the correct values for benefits and costs. This result is exactly what one would expect in light of the discussion in Unit 7. There, we saw that competitive markets are efficient. In such markets, the price of the good or service is an exact measure of its marginal benefit to consumers and its marginal cost to producers: P _MB _MC. For instance, if construction of the bridge requires 50,000 cubic yards of concrete and the price of concrete is $100 per cubic yard, the total cost of this input is $5 million. The same principle applies to the cost of any input to production—capital, labor, land, and so on. It also applies to valuing the benefits of program outputs. For instance, suppose the chief benefit of constructing a water project is the irrigation of new tracts of land. The market value of water would represent the dollar benefits of the project.1

Nonmarketed Benefits and Costs

One gains a renewed appreciation for the role of market prices when one considers the difficulties in valuing nonmarketed items. For instance, how can we judge the benefits of public schools? Should performance be judged by average test scores? dropout rates? Indeed, educators agree there are no ideal performance measures. Because public education is provided collectively (i.e., financed out of local tax revenues), there is no "market" value for this essential service. Parents do not pay market prices for their children's education. Contrast these difficulties with the problem of valuing education provided by private schools. Here the value is clear; it is at least as much as the price parents actually pay in tuition. There is no need to study the determinants of school performance; the market price is enough. If a private school fails to deliver a quality education, parents will stop paying the high market price.

SOCIALLY DETERMINED VALUES

Society, via its norms and laws, places monetary values on many nonmarketed items. Workers' compensation laws determine monetary payments in the event of industrial injuries. Judges and juries determine the extent of damages and appropriate compensation in contract and tort proceedings. In divorce cases, the court frequently is asked to determine the monetary value of a homemaker's contribution to the family. Government regulations implicitly determine societal values. For instance, federal

law requires special access for the handicapped in public buildings and public transit. Presumably, the cost of meeting this requirement represents a lower bound for the value society places on easy access.

VALUING LIVES

Perhaps the most controversial application of benefit-cost analysis occurs in the valuation of lives. Many of us would like to believe human life is priceless and beyond monetary measure. Yet a host of government programs involves determining whether enhanced safety, not only in the form of injuries prevented but also in terms of lives saved, is worth the cost. As mentioned earlier, the decision not to spend $240 million on a program expected to save 50 lives means that the value of a life is less than $4.8 million, the implied cost per life. In short, spending or not spending on public safety programs implicitly or explicitly involves valuing lives.

A number of approaches to estimating the dollar value of a life have been taken. Although none can produce a definitive dollar value, the methods do target a likely value range. A first approach, the earnings method, appeals to the labor market for an answer. Boldly stated, the value of a life is measured by the present value of an individual's lifetime wage earnings. Depending on the precise assumptions, studies that have used the earnings approach have produced estimates in the range of $3 million to $4 million per life. Of course, many would argue that the presumption "you're worth what you earn" constitutes a gross understatement of a life's value. (One would never want to apply this method to unemployed or retired people.)

A second approach examines the amounts of compensation individuals demand for bearing the risk of death. Other things being equal, high-risk jobs—law enforcement, fire fighting, skyscraper construction, mining, lumber jacking, oil drilling, to name a few—pay higher wages. The wage premium that people demand for taking on a greater risk of death, gives us an idea of how they value increased or decreased risk of death. For example, for skyscraper construction workers, the additional mortality risk is approximately .2 percent per year. Suppose the wage premium paid to such workers (again relative to a comparable low-risk job) is $12,600 per year. What conclusions can we draw from these facts? If a construction firm hires, say, 1,000 workers, it pays a total wage premium (due to risk) of $12.6 million and

2 deaths will occur on average. The implied value of a life is 12.6/2 _ $6.3 million.

Proponents argue that the risk-cost trade-off embodied in private markets is the best guide to valuing lives when it comes to government decisions. However, for several reasons, the method is likely to underestimate a life's dollar value. An individual who chooses a high-risk occupation is likely to be more risk loving than the average person and, therefore, demand a lower wage premium. (If compensation for the average person were closer to $15,000, the value of a life would be $7.5 million.) In addition, workers in dangerous occupations may be inadequately informed about the true risks. Values for lives inferred from such decisions may reflect (at least partially) poor judgment, as well as

calculated risks. Also, many high-risk jobs may go to people who, due to their socioeconomic status, have few other options.

U.S. regulatory agencies use different dollar amounts for valuing lives, ranging from $5 million at the Food and Drug Administration (FDA) to almost $9 million at the Environmental Protection Agency. While there is no single correct value, it's important to remember that higher values mean higher benefits (relative to costs) and so tip the scales toward a greater degree of safety related regulation.

REVIEW QUESTION

1. Discuss about market failures and regulation related to it.
2. Describe market failure due to monopoly.
3. Discuss about market failure due to externalities.
4. Discuss about remedying externalities.
5. Describe private payments.
6. What is the process of promoting positive externalities?
7. Discuss about market failure due to imperfect.
8. Discuss about benefit-cost analysis and public goods.
9. What are the basics of benefit-cost analysis?
10. Describe the process of evaluating a public project.

FURTHER READINGS

1. Managerial Economics - Arun Kumar, Rachana Sharma
2. Managerial Economics - Thomas J. Webster
3. Managerial Economics - Petersen / Jain
4. Managerial Economics - Yogesh Maheshwari
5. Managerial Economics - E. Narayanan Nadar, S. Vijayan

UNIT-12 DECISION MAKING UNDER UNCERTAINTY

CONTENTS

INTRODUCTION

In this Unit, we begin our study of decision making under uncertainty. First, we review the fundamentals of uncertainty, probability, and expected value. Then we examine the use of decision trees as a guide for managerial choices, especially in sequential decisions. Finally, we explore the effect of risk aversion on managerial decisions: how a manager can assess attitudes toward risk and apply the expected-utility criterion as a decision guide.

UNCERTAINTY, PROBABILITY, AND EXPECTED VALUE

Uncertainty lies at the heart of many important decisions. Managers are often uncertain about outcomes that have a direct bearing on the firm's profit. For example, introducing a new product entails a multitude of risks, including the cost and timetable of development, the volume of sales in the product's first and subsequent years, and competitors' possible reactions. The example that opens this Unit suggests that uncertainty concerning the future course of the macro economy—consumer and business spending, price inflation, interest rate movements—is an important factor for many industries and firms. Uncertainty (or risk) is present when there is more than one possible outcome for a decision.1 Roughly speaking, the greater the dispersion of possible outcomes, the higher the degree of uncertainty. The key to sound decision making under uncertainty is to recognize the range of possible outcomes and assess the likelihood of their occurrence. Uncertainty is acknowledged in expressions such as "it is likely," "the odds are," and "there is an outside chance." The difficulty with such qualitative expressions is that they are ambiguous and open to different interpretations. One is prompted to ask, "How likely is likely?" The essential means for quantifying statements of likelihood is to use probabilities. It is far more useful for a meteorologist to state that there is a 60 percent chance of rain tomorrow than to claim that rain is likely. Probability has been described as the mathematical language of uncertainty. The key is to have a sound understanding of what probabilities mean. The probability of an outcome is the odds or chance that the outcome will occur. In the usual parlance, we speak of probabilities as ranging between 0 and 1. (An event having a probability of 1 is a certainty; an event having a probability of 0 is deemed impossible.) Whatever the probability, the relevant question

is: What is the basis for this assessment? Frequently there is an objective foundation for the probability assessment. The chance of heads on a single toss of a fair coin is 50 percent, or one-half. In a random draw, the chance of picking the lone black ball from a hat containing five balls is one in five, and so on. When viewed closely, the main basis for assessments such as these is the notion of a probability as a long-run frequency. If an uncertain event (like a coin toss or a random draw) is repeated a very large number of times, the frequency of the event is a measure of its true probability. For instance, if a fair coin is tossed 1,000 times, the frequency of heads (i.e., the number of heads divided by the total number of tosses) will be very close to .5. If the actual long-run frequency turned out to be .6, we would be justified in asserting that the coin was unfair. The frequency interpretation applies to most statistical data. For example, if annual employment in the mining industry totals 40,000 workers and

80 workers die in mining accidents each year, the annual probability of a representative mine worker dying on the job is 80/40,000 or .2 percent. It should be evident that in many (and perhaps most) situations, there is no chance that a situation will be repeated and therefore no way to assess probabilities on frequency grounds. In its development of a new product (one that is unique to the marketplace), a firm knows that the product launch is a one-shot situation. The firm may believe there is a 40 percent chance of success, but there is no way to validate this by launching the product 100 times and watching for 40 successes. Similarly, a company about to enter into patent litigation faces the problem of predicting the likely outcome of a one-time legal suit. Still another example is a business economist attempting to put odds on the likelihood of a new oil price "shock" (say, a 50 percent rise in oil prices) over the next 18 months. In dealing with such situations, decision makers rely on a subjective notion of probability. According to the subjective view, the probability of an outcome represents the decision maker's degree of belief that the outcome will occur. This is exactly the meaning of a statement such as "The chance of a successful product launch is 60 percent." Of course, in making a probability assessment, the manager should attempt to analyze and interpret all pertinent evidence and information that might bear on the outcome in question.2 For the new product, this would include consumer surveys, test-market results, the product's unique qualities, its price relative to prices of competing products, and so on. The point is that a subjective probability is not arbitrary or ad hoc; it simply represents the decision maker's best assessment, based on current information, of the likelihood of an uncertain event. In this sense, all probabilities—even those based on frequencies or statistical data—represent the decision maker's degree of belief.

Expected Value

The manager typically begins the process of analyzing a decision under uncertainty by using a probability distribution. A probability distribution is a listing of the possible outcomes concerning an unknown event and their respective probabilities. As we saw earlier, assessing relevant probability distributions is the first step in the manager's analysis. For

example, the manager might envision the probability distribution shown in the table for the first year's outcome of a new-product launch.

DECISION TREES

The decision tree is a convenient way to represent decisions, chance events, and possible outcomes in choices under risk and uncertainty. In fact, this simple diagram can incorporate all the information needed to "solve" the decision problem once the specific objectives of the decision maker have been established. The method is extremely versatile. When first encountered, choices under risk appear messy, ill defined, and puzzling. The actual choices, the potential risks, and the appropriate objective to pursue may all be far from clear. The individual should not be blamed for regarding his or her choice as "a riddle wrapped in a mystery inside an enigma," to borrow a phrase from Winston Churchill. However, sketching a crude decision tree almost always will clarify the options. The very structure of the tree emphasizes the ingredients (choices, outcomes, and probabilities) necessary for making an informed decision. The more precise the tree becomes (after drawing and redrawing), the more precise one's thinking becomes about the problem. The "finished" tree can then be evaluated to "solve" the decision problem. Probably more important, the decision tree provides a visual explanation for the recommended choice. One easily can pinpoint the "why" of the decision: which circumstances or risks weighed in favor of which course of action. And one can undertake any number of sensitivity analyses, altering the facts of the decision to determine the impact on the recommended course of action.

Decision trees can be simple or complex, spare or "bushy," small enough to evaluate by hand or large enough to require a computer. To illustrate the method, we start with a concise example.

Features of the Expected-Value Criterion

The depiction of the risk in 12.1 hardly could be simpler. Thus, it comes as no surprise that the expected-value calculation is automatic, indeed, almost trivial. Nonetheless, it is important to recognize the general properties of this criterion, properties that apply equally to simple and complex risks.

The first (and most basic) feature of the expected-value standard is that it values a risky prospect by accounting not only for the set of possible outcomes, but also for the probabilities of those outcomes. For instance, suppose the wildcatter must decide whether to drill on one site or another. (There are insufficient resources to drill on both.) The first site's possible monetary outcomes are 800, 600, 160, _60, and _200 (all in thousands of dollars); these outcomes occur with probabilities .05, .15, .2, .25, and .35, respectively. Thus, the expected profit from drilling this site is (.05)(800) _ (.15)(600) _ (.2)(160) _ (.25)(_60) _ (.35)(_200), or $77 thousand. The second site has the same five possible outcomes as the first but with probabilities .05, .2, .25, .2, and .3. Notice that the second site offers higher probabilities of "good" outcomes than the first site. Clearly, then, the second site should have a higher value than the first. The expected-value standard satisfies this common-sense requirement. Performing the appropriate computation will show that the second site's expected profit is $128,000, a significantly higher than the expected

profit of the first site. Second, the expected value of a risky prospect represents the average monetary Outc me if it were repeated indefinitely (with each repeated outcome generated independently of the others). In this statistical sense, the expected-value standard is appropriate for playing the long-run averages. Indeed, many managers employ the expected-value criterion when it comes to often-repeated, routine decisions involving (individually) small risks. For instance, suppose you have the chance to bet on each of 100 tosses of a coin. You win a dime on each head and lose a nickel on each tail. This, you'll no doubt agree, is the epitome of a routine, often-repeated, low-risk decision. Here the expected-value criterion instructs you to bet on each toss. If you choose this profitable (albeit somewhat boring) course of action, your expected gain in the 100 tosses is $2.50. Your actual profit will vary in the neighborhood of $2.50, perhaps coming out a little above, perhaps a little below. The statistical "law of large numbers" applied to the independent tosses ensures that there is no real risk associated with the bet. Third, in decisions involving multiple and related risks, the expected-value criterion allows the decision maker to compute expected values in stages. 12.2 makes this point by presenting a "bushier" (and more realistic) tree for the wildcatter's drilling decision. The tree incorporates three risks affecting drilling profits: the cost of drilling and recovery, the amount of oil discovered, and the price of oil per barrel. As the tree depicts, the cost of drilling and recovery is the first uncertainty to be resolved and depends on the depth at which oil is found (or not found). In the wildcatter's judgment, oil may be struck at one of two d epths or not at all. Thus, the tree depicts three branches emanating from the initial chance node. As an example, let's consider the second branch: oil found at 5,000 feet. This branch ends in a chance node from which three new branches emerge. These branches show the possible amounts of oil (barrels per year) that might be recovered; the third branch, for instance, has a total recovery of 16,000 barrels. Finally, each recovery branch ends in a chance node from which three new branches sprout. These indicate the possible different values of average oil prices over the life of the well. For example, the third branch lists a $55-per-barrel price. At the end of this branch, the last uncertainty is resolved and the wildcatter's profit, in this case $180 thousand, is finally determined. (Simply take the profit s at face value. We have not supplied the revenues and costs on which they are based.) The path from the leftmost chance node to the $180,000 profit outcome indicates one particular scenario that might occur: finding a 16,000-barrel oil field at 5,000 feet and selling it at a two-year average price of $55 per barrel. However, this outcome is but one of many possible outcomes contingent on the resolution of the multiple risks. In all there are (2)(3)(3) _ 1 _ 19 possible profit outcomes, one for each branch tip. The combination of multiple risks, each with multiple outcomes, means that the corresponding decision tree will be bushy indeed.

The bushy tree also requires a lengthier process of probability assessment, because the wildcatter must evaluate probabilities for three distinct risks. The first three branches of the tree show his chances of striking (or not striking) oil at different depths. If he finds some oil at a

given depth, the next question is how much. The secondary branches of the tree list the chances of finding different oil quantities. Note that the likelihood of different recovery amounts depends on the depth at which oil is first found, and the likelihood of very large deposits is better at 3,000 feet than at 5,000 feet. (Remember that these recovery probabilities are conditional on some oil being found at all. Shallow fields are likely to be large fields, but the chance of finding oil at 3,000 feet is only .13 in the first place.) Finally, once the recovery quantity is ascertained, the sole remaining uncertainty concerns the market price of oil. The chances listed on the third-level branches have been obtained from an expert's prediction of future prices. Note that the chances of different market prices per barrel are independent of the quantity of oil recovered (i.e., the chances are the same regardless of the recovery amount).

A More Complicated Drilling Decision This decision tree contains multiple risks that generate 19 possible outcomes. reserve at a depth of 3,000 feet. The three branches list the profit outcomes for this field depending on the (uncertain) oil price. The expected profit from such a field is simply the average of the possible profit outcomes weighted by the respective probabilities. Thus, the expected profit is (.2)(700) _ (.5)(350) _ (.3)(150) _ $360 thousand, listed in chance node D. But what if the field had yielded 8,000 barrels per year? By an analogous calculation, we find the expected profit to be $636 thousand in this case, as shown in chance node E. The expected profits for the chance nodes F through I (corresponding to different sized fields at different depths) have also been computed and listed on the tree.

At this point, we have "averaged out" the price uncertainty. In the next step, we average over the possible quantities of oil found. Chance node B shows the expected profit if oil is found at 3,000 feet, computed by averaging the expected profits at nodes D through F:

ECONOMIC CONDITIONS

The 1990s recession in Europe and the late 1990s financial crisis in Southeast Asia caused dramatic falls in business and consumer spending. Global firms with sales concentrated in these regions saw profits evaporate and losses mount.

UNCERTAIN COSTS

Because of low-skilled workforces, lack of capital, and primitive distribution systems, the costs of doing business in developing countries are frequently high and uncertain. Foreign firms assembling electronics goods in Russia have been plagued by low worker productivity, vandalism, and crime.

DIFFERENT CULTURES

Brazilians spend a higher percentage of income on their children than do citizens of neighbouring countries. They are eager for disposable diapers, while Argentines are largely indifferent. Consumers in Southeast Asia are accustomed to buying light meals from street vendors, not from fast-food restaurants. To cite an extreme case of cultural miscalculation, General Motors introduced its popular Nova car model into South

America. Only after disastrous sales did the company realize that no va means "does not go" in Spanish.

POLITICAL RISK

Tax and regulatory burdens, government bureaucracy and even corruption, and changing political parties and governments all contribute to the risk of doing business abroad. Over the past 50 years, international businesses have been decimated by unrest and civil war in places such as Cuba, Lebanon, El Salvador, Vietnam, and the Balkans. Today, outright expropriation is much less frequent but remains a risk.

EXCHANGE-RATE RISK

A firm that earns a significant part of its revenues abroad is subject to exchange-rate risk when converting these to its home currency. For instance, a depreciating Japanese yen means lower dollar profits from revenues earned in Japan. Similarly, the costs incurred by a foreign subsidiary are subject to exchange-rate risk. Thus, the depreciating currencies of Southeast Asia (by lowering the dollar-equivalent costs) make production in that part of the world more attractive to global firms.

SEQUENTIAL DECISIONS

Some of the most interesting and important business and economic problems call for a sequence of decisions to be made over time. For example, suppose a chemical firm is considering a large capital investment in a new petrochemical facility. The profitability of such an investment depends on numerous uncertain factors: future market demand, reactions of close competitors, and so on. Profits also depend on the future product and pricing decisions of the firm itself. It is not simply that the firm faces many decisions over time; the more important point is that the sequence of decisions is interdependent. A correct investment decision today presupposes that the company will make optimal (i.e., profit-maximizing) pricing decisions tomorrow if the plant is built. The following example illustrates this general point about sequential decisions.

Research shows that individuals have difficulties identifying and evaluating risks. Too often they rely on intuition, rules of thumb, and experience to make risky decisions. Managers' most common pitfalls include: 1. Seeing too few possibilities. That is, they take a too narrow or "myopic" view of the future. While successful firms astutely foresee possible future consequences and act appropriately, many firms suffer losses by failing to foresee coming events. Too often managers simply extrapolate the current status quo into their forecasts for the future, thus ignoring upside and downside possibilities alike. Professor Max Bazerman of Harvard Business School calls these risks "predictable surprises" (the disasters you should have seen coming). It is like drawing a decision tree with whole sections of chance branches missing (because those possibilities have been overlooked) but not knowing it.

2. Relying on verbal expressions of probability. Losing the patent case is unlikely. There is a reasonable chance that our product will beat our rival's to market. Although expressions such as these come naturally, they are surprisingly imprecise. Researchers have asked scores of individuals, including businesspeople, what a host of such expressions

mean in terms of probability. For instance, "unlikely" conveys a probability of anywhere between 11 and 39 percent, with a median response of 25 percent. In turn, "a reasonable chance" can mean a probability as high as 80 percent or as low as 50 percent. As decision trees remind us, determining reasonable probabilities for the risks that matter is crucial for crafting profit-maximizing decisions. A pessimistic view of "unlikely" could well lead to a very different decision than an optimistic view. It's far better to try to pinpoint and agree on reasonable probability estimates in the first place.

3. Holding optimistic beliefs. Here, optimism means overstating the probability of favorable outcomes and downplaying the chances of unfavorable ones. By nature confident, many mangers unconsciously engage in wishful forecasting: What they want to have happen they believe is likely to happen. Clearly, overoptimistic, unrealistic beliefs can lead to poor or even disastrous decisions. A constructive remedy to unfounded optimism is to insist on realistic assessments based on external benchmarks. Nobel prize winner Daniel Kahneman calls this "taking the outside view." For instance, a management team might believe and claim a 60 percent chance of success for a new product, based purely on an internally focused assessment. But of all new product launches surveyed each year, only about 10 to 15 percent are successfully being sold two years later. Even if impressive internal factors justify elevating the product's success rate to, say, three times this base rate, a realistic revised probability is only 30 to 45 percent—a far cry from 60 percent.

RISK AVERSION

Thus far, we have used the concept of expected monetary value as a guide to making decisions under uncertainty. A decision maker who follows the expected-profit criterion is said to be risk neutral. This standard is appropriate for a manager who is willing to play the averages. The evidence suggests, however, that individuals and firms are not neutral toward risks that are large relative to their financial resources. When it comes to significant risks, individuals and institutions adopt an attitude that is conservative toward losses. Thus, the use of the expected-profit criterion must be qualified.

A COIN GAMBLE You are offered the following choice: You can receive $60 for certain (the money is yours to keep), or you can accept the following gamble. A fair coin is tossed. If heads come up, you win $400; if tails come up, you lose $200. Would you choose the sure $60 or accept the gamble on the coin toss? In answering, imagine that real money (your own) is at stake. When given this choice, the majority of individuals prefer the sure $60 to the gamble. This is not surprising given the magnitude of the risk associated with the coin toss. Notice, however, that choosing $60 is at odds with maximizing expected profit. The expected profit of the coin toss is (.5)(400) _ (.5)(_200) _ $100. Thus, a risk-neutral decision maker would prefer the gamble to the sure $60. Refusing the bet shows that you are not risk neutral when it comes to profits and losses of this magnitude.

A precise way to express one's evaluation of the coin toss (or any risky prospect) is to name a certainty equivalent. The certainty equivalent (CE)

is the amount of money for certain that makes the individual exactly indifferent to the risky prospect. Suppose that, after some thought, you determine you would be indifferent to the options of receiving $25 for certain or facing the risk of the coin toss. You are saying that your CE for the coin toss is $25. This CE is significantly smaller than the expected value of the bet, $100. This being the case, we would say that you are risk averse. An individual is risk averse if his or her certainty equivalent for a given risky prospect is less than its expected value. Loosely speaking, the magnitude of one's aversion to risk is indicated by the shortfall of the CE below the expected value of the risky prospect; this difference (sometimes referred to as a discount for risk) measures the reduction in value (below expected value) due to a prospect's riskiness. Here the risk discount is 100 _ 25 _ $75. The discount depends on individual preferences as well as on the size of the risk. For instance, a second individual might prefer to avoid the coin toss altogether; that is, in a choice between the coin toss and receiving $0 for certain, this individual prefers $0. This preference makes good sense for someone who does not wish to bear the downside risk of the coin toss. Suppose this individual is indifferent to the options of paying $20 for certain or taking the coin toss. (He or she is willing to pay $20 to avoid the risk of the gamble.) Here the CE is _$20, and the risk discount is $100 _ (_20) _ $120. Clearly, the second decision maker is more risk averse than the first.

THE DEMAND FOR INSURANCE

Risk aversion provides a ready explanation concerning the demand for insurance. Insurance companies stand ready to compensate their policyholders in the event of losses (specified in the insurance contract) at a price in the form of the premium paid by the customer to the company. Risk-averse individuals are willing to give up monetary income to avoid risks. In effect, this is what they do when they purchase insurance. To make the argument concrete, consider a couple who is about to purchase fire insurance to protect their home (which is valued at $150,000). The risk of a fire destroying their house is very small—about 1 in 300 in any given year. Nevertheless, the loss of their house would mean financial ruin. Thus, the couple finds it prudent to purchase insurance. In return for payment of a $500 annual premium, a 100 percent fire policy promises to pay whatever amount is necessary to rebuild and replace the house in the event of fire. In purely financial terms, the couple faces the following options. If they do not buy the policy, their wealth at the end of the year will be $150,000 if there is no fire or $0 if a fire occurs (a 1-in-300 chance). Their expected wealth is $149,500. (Check this for yourself.) By purchasing the policy, their net wealth is $150,000 _ $500 _ $149,500 at the end of the year. Their wealth is certain. Regardless of whether a fire occurs, they will have their house (or the money to rebuild it). Notice that whether or not they purchase insurance, the couple's expected wealth is the same, $149,500. Because they are risk averse, the couple prefers the certain $149,500 provided by insurance to the alternative of bearing the risk of fire. Thus, they purchase full insurance.

In this example, the company has offered the couple "actuarially fair" insurance; that is, the couple's premium ($500) just covers the company's expected payout under the policy: (1/300)($150,000) _ $500. Because of their large size and ability to pool different risks, insurance companies generally behave as though they are risk neutral. To illustrate, suppose the company insures 300,000 houses in a state against fire. Although it is impossible to predict which houses will be struck by fire, the law of large numbers indicates that very close to 1,000 homes in total will have fire losses. Thus, the total premiums ($150 million) will closely match the company's actual payout. Because of administrative costs in writing the policies, insurance companies typically charge premiums that exceed their expected losses. (Of course, competition among insurance companies limits the premiums any one company can charge.) But higher premiums do not eliminate (although they may reduce) the demand for insurance. Even if the fire insurance premium were $1,000 per year, the risk-averse couple might leap at the chance to buy coverage rather than go unprotected.

Expected Utility

How can a manager formulate a criterion, reflecting the firm's attitude toward risk, to guide his or her decisions? The formal answer to this question was developed more than 50 years ago by mathematical economists John Von Neumann and Oscar Morgenstern, and is called the expected-utility rule. (At the same time, Von Neumann and Morgenstern developed the field of game theory, which we encountered in Unit 10.) The use of expected utility proceeds in two steps. First, the decision maker must think carefully about the firm's preferences concerning risks: what risks it is willing to accept and how to value those risks. In the process, the manager constructs a utility scale that describes this risk tolerance. Second, the manager analyzes the decision problem in much the same way as before, that is, he constructs

a decision tree showing relevant probabilities and possible monetary outcomes and then evaluates the tree. However, there is one crucial difference: In contrast to the risk-neutral manager, who averages monetary values at each step, the risk-averse decision maker averages the utilities associated with monetary values. At each point of decision, the manager selects the alternative that supplies the maximum expected utility. With this summary in hand, let's see exactly how the method works.

A RISK-AVERSE WILDCATTER

Once again, let's consider the wildcatter's basic decision problem, reproduced in 12.7. Now suppose the wildcatter is risk averse; he is unwilling to rely on expected profits as his choice criterion. Instead, he seeks to determine a criterion for choosing among risky prospects that reflects his own attitude toward risk. We now show how he can construct a utility function that measures his own degree of risk aversion and how he can use this function to guide his choices.

The wildcatter begins by attaching a utility value to each possible monetary outcome. Let's start with the decision to drill. Here the outcomes are $600,000 and -$200,000; these are the best and worst possible outcomes, respectively. The wildcatter is free to set these utility

values arbitrarily as long as the best outcome receives the higher value. The usual choice is to assign the worst outcome a utility value of zero. Thus, we would write U(_200) _ 0; that is, the utility associated with a loss of $200,000 is zero. In turn, let's arbitrarily set U(600) _ 100. This establishes a range of utility values from 0 to 100 for monetary outcomes between the worst and best possible outcomes.

Using these utility values, the wildcatter evaluates the option to drill by computing its expected utility. The expected utility is the probability of each outcome times its utility, summed over all outcomes. Thus, the expected utility of drilling is Now consider the "do not drill" option. In this case, the wildcatter's monetary result is $0 for certain. What utility value should the wildcatter assign this outcome? To determine U(0), the wildcatter compares $0 for certain with a gamble offering $600,000 (with probability p) and _$200,000 (with probability 1 _ p).The wildcatter measures his relative preference for $0 by finding the probability, p, that leaves him indifferent to the options of $0 and the gamble. Suppose that, after some mental trial and error, he judges his indifference probability to be p _ .5;

REVIEW QUESTION

1. Describe decision trees in detail.
2. What is risk aversion?
3. What is the demand for insurance?
4. Describe the risk-averse wildcatter.

FURTHER READINGS

1. Managerial Economics - Arun Kumar, Rachana Sharma
2. Managerial Economics - Thomas J. Webster
3. Managerial Economics - Petersen / Jain
4. Managerial Economics - Yogesh Maheshwari
5. Managerial Economics - E. Narayanan Nadar, S. Vijayan

UNIT-13 ASYMMETRIC INFORMATIONAND ORGANIZATIONAL DESIGN

CONTENTS

INTRODUCTION

In Units 12 and 13, we considered decision making under uncertainty and the value of information solely from the individual manager's point of view. Dealing with risk and acquiring better information are equally relevant in decisions involving multiple decision makers, whether they operate within markets, across organizations, or within the same organization. Particularly important examples involve asymmetric information—situations in which one party knows more than another about key economic facts. As we shall see, the presence of asymmetric information has a number of implications. First, managers must be careful to draw correct inferences from the behavior of others. Second, asymmetric information can lead to market failures, that is, can impede profitable transactions between buyers and sellers. Third, it can create incentive problems. One party may undertake behavior that is not in the best interest of another party with which it interacts. The first section of this Unit considers the effects of asymmetric information, and the second explores how firms can best organize themselves to deal with asymmetric information.

ASYMMETRIC INFORMATION

In situations characterized by asymmetric information, one party knows more than another about key economic facts. The presence of asymmetric information can lead to adverse selection and moral hazard, each of which we take up in turn. Adverse Selection As noted in Unit 13, managers must make accurate probability assessments in order to make well-informed decisions. But as the next example shows, these assessments must take into account the likely behavior of other decision makers.

A BENEFITS PROGRAM After considerable planning, a company's human resources department has introduced a premium medical insurance program 582 Unit 14 Asymmetric Information and Organizational Design for its employees and their spouses. Employees who elect this coverage pay more than with the standard plan. Among other benefits, the premium plan will pay for maternity-related health expenses. The firm estimates that 1 in 20 of its employees will have a new baby in a given year. (This estimate comes from records for the last 10 years.) Accordingly, the company has set the premium to cover its expected payouts at this 1-in-20 rate. Postscript: In the first two years of the program, the company has lost an enormous amount of money on the program. Employees covered by the plan are having babies at the rate of 1 in 10 per year. Is this bad luck or bad planning?

The company's losses are not due to bad luck. Today's workforce does not differ in its composition or behavior from that of the last 10 years. Instead, the firm's losses are due to adverse selection. The following table lists the hypothetical, but plausible, numbers for the first year of the program. Notice that the overall rate of new babies is 200/4,000 or 1 in 20, exactly the average rate of the previous 10 years. The rate of having babies has not changed. However, among policyholders, the rate of having babies is 1 in 10 (100/1,000); among non policy holders, it is 1 in 30. This result should not surprise us. Couples who are planning to add to their families will tend to elect the policy; those who are not will forgo the coverage. This behavior usually is termed self-selection. From the company's point of view, the result is called adverse selection. Couples who are most likely to have babies (and know this) will most likely elect the coverage.

Adverse selection occurs because of asymmetric information. Individuals have better information about their true risks than the insurance provider does. As a result, individuals at the greatest risk elect insurance coverage. To avoid losses on their policies, insurance companies must anticipate this behaviour and set their premiums accordingly. In the preceding example, the company would have to double its premium to break even.

A "LEMONS" MARKET The used-car market is a famous example of asymmetric information.2 Consider someone trying to sell a car that is six months old and has been driven only 4,000 miles. Even though almost new, it now may sell for as little as 75 percent of its original sale price. The steep discount 2The classic article on this topic is G. A. Akerlof, "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," Quarterly Journal of Economics (1970): 488–500. comes from adverse selection. A typical buyer of used cars considers, "What kinds of used cars would sellers most likely offer for sale?" The answer is low quality cars, including the lemons. The fact that the car is for sale should make the buyer suspicious.

Signalling

In the presence of asymmetric information, managers gather information to better gauge risks. For instance, auto insurers place drivers into different risk classes according to past driving record (as well as age and gender) and set premiums accordingly. A used-car buyer might have a

licensed mechanic thoroughly check out a prospective purchase. Banks and other lenders devote significant time and resources to assessing borrowers' computerized credit histories. By obtaining better information (albeit at a cost), the manager can go a long way toward reducing or even eliminating the problem of adverse selection.

Asymmetric information poses a problem for the informed party as well. For instance, a seller may know he or she has a high-quality used car but may be unable to sell it for its true value due to adverse selection. Similarly, an individual who cannot prove he or she is a good credit risk may have to pay the same (high) interest rates as high-risk candidates. In general, "high-quality" individuals wish to distinguish themselves from their "low-quality" counterparts. They can do this in several ways. One way is by developing a reputation. For example, a seller may seek to build and maintain a reputation for delivering high-quality goods and services. A business that depends on repeat purchases and word of mouth will find its long-term interest served by accurately representing the quality of its goods.

An alternative method is to offer guarantees or warranties. A warranty serves two related purposes. First, it protects customers against quality problems or defects in the products. Second, the warranty offer itself signals product quality. A producer of a high-quality product can afford to offer an extensive warranty because guaranteeing a reliable product will cost the producer very little. A producer of a low-quality product will choose not to offer such a warranty because it is much more expensive for it to do so. In short, the offer of a warranty provides a (low-cost) way for high-quality producers to distinguish themselves from low-quality producers. Signaling quality in this way allows high-quality producers to charge higher prices for their goods and services. Of course, warranties may not always produce the desired result. In 2002 Amtrak withdrew a warranty whereby dissatisfied customers were entitled to free future train travel. Amtrak found itself issuing more and more free travel for situations that it could not control, including weather delays and delays due to having to cede priority to freight trains.

Signaling is a common response to the presence of asymmetric information. A particularly important example occurs in job markets. At the time of hiring, a firm may find it difficult to predict how productive different job candidates will be. Certainly, management will have much better information after the worker has been on the job for six months or a year, but by that time, management may have invested considerable resources in on-the-job training for the worker and may have little flexibility in modifying its decisions. (Terminating unproductive workers is difficult and costly.) If the firm cannot distinguish between high- and low-quality workers at the time of hiring, the best it can do is offer the same wage (based on average productivity) to all new workers. (Low-quality workers are paid more than their worth and high-quality workers less than their worth.) But the workers themselves are well aware of their abilities, skills, and energy. High-productivity workers would like to signal their true abilities to potential employers and thereby obtain higher-paying jobs. One way to signal their true value is via education.5 Thus, education not only provides knowledge—ways of thinking as well

as specific skills—that can increase an individual's productivity, it also serves as a signal. Even if it did not contribute to productivity, education would continue to signal productivity. Individuals of high innate ability find school easier and perform better. Thus productive persons have greater benefits and lower costs from additional years of education and will invest more heavily in education than will their less productive counterparts. Other things being equal, the higher an individual's educational achievement (measured by years of schooling, advanced degrees, and so on), then the greater is his or her potential productivity.

Health Insurance and Medical Costs forward to future years. Thus workers, in their medical and health care decisions, should begin acting as if they are spending their own money, at least up to the amount of the deductible.

An alternative approach targets physician incentives. The traditional fee-for service approach simply reimburses doctors for the cost of treatment and gives neither the patient nor the doctor an incentive to consider costs. When ordering outside tests, doctors are often unaware of the cost of the tests. Even worse, in some practice areas, doctors have a financial incentive to prescribe costly treatments, even if less costly treatments would be more effective. Moreover, under the current legal regime, the threat of malpractice suits encourages doctors to practice defensive medicine—to overprescribe costly tests and treatments. In contrast to a fee-for-service approach, fixed payments for treatments, separated into diagnosis-related groups, provide strong cost incentives. For instance, a hospital that receives a fixed payment for a surgical appendectomy has a strong incentive to keep down costs. If its cost exceeds the fixed payment, it bears the loss; if its cost is below the fixed payment, it garners the profit. An increasingly popular payment scheme, the capitation approach, takes incentives to the limit. Under capitation, an HMO pays a group of doctors one fixed annual payment per patient. The fee per patient is set at the estimated cost of caring for each enrollee. At the end of the year, if total costs come in below total fees, the doctors pocket the profit. Conversely, if costs outpace fees, the doctors absorb the loss. However, whether such incentive plans will always lead consumers to wiser health decisions is an open question. A 2004 Rand Corporation study found that worker co-payments reduced the use of prescription drugs but increased visits to emergency rooms, raising the concern that patients, made to pay for a significant share of their health costs, were not buying the medicines they needed, leading to more expensive treatment in the future. Other research has found that many older women were foregoing mammograms to detect breast cancer, in spite of government guidelines recommending regular mammograms.

The Choice between In-House Production and Outsourcing

Factors Favoring In-House Production	Factors Favoring Outsourcing
1. Firm-specific good or service	1. Standardized good or service
2. Outside risks: input quality, supply disruptions	2. Competitive market available
3. High degree of coordination required	3. Low degree of coordination required

A corollary to this proposition follows. Organizations should distribute tasks to best generate and utilize specialized information. This is not a new insight. After all, division of labor and specialization characterize the modern firm. However, specialization not only enhances productivity in the traditional sense but also greatly improves the quality of decisions. Imagine the following nightmare. You are a top executive whose daily calendar calls for you to make six crucial decisions: from solving a production problem in your West Coast plant to deciding on a new marketing plan, to dealing with federal regulators. You are woefully unprepared to decide any of these issues. (This is sort of like the exam no one told you about and for which you never studied.) By comparison, suppose you are an executive waking up to face six big decisions in your bread-and-butter area of responsibility. Well prepared by years of experience and accumulated knowledge, you eagerly tackle these challenges. Modern firms typically divide responsibilities along functional lines—production, marketing, finance, and so on. This type of structure has obvious advantages and less obvious disadvantages. One risk is that functional managers may lose sight of the bigger picture. Obviously, a materials manager must communicate with a production manager. The latter cannot plan to increase jeans production without the necessary denim and thread on order. Similarly, a manufacturing manager greatly benefits by learning of customers' needs and complaints from the marketing manager.

An alternative organizational design divides responsibility by line of product or service. Product lines represent natural profit centers. Consequently, a product manager would oversee many functional areas for his or her product and make decisions to maximize profitability. Midlevel managers would still occupy functional jobs within this product division. Management can also be organized by the type of customer (business versus residential, for instance) or by geographical regions. Some corporations choose to combine the advantages of product and functional divisions. For example, the marketing managers from the firm's several divisions might meet periodically as a team for coordination and consultation.

The Airbus Industries consortium was formed to produce commercial aircraft in 1970.16 Backed financially by its four European parents, the commercial aircraft manufacturer lived the first 30 years of its existence as a consortium of French, British, German, and Spanish aerospace companies, plus a marketing unit charged with negotiating the sale of aircraft.

DECENTRALIZATION

The trend toward dividing organizations along functional and product lines creates pockets of specialized information dispersed throughout the business organization. This phenomenon virtually precludes the possibility of completely centralized decisions by top executives and the board of directors. As a practical matter, even the most engaged CEO can make only a small fraction of the decisions involved in managing a modern business. To an increasing degree, the norm in the modern corporation is decentralization.

Decentralization assigns decision-making responsibilities as closely as possible to the holders of relevant information. Presumably, an experienced regional sales manager with his or her ear close to the ground can best make periodic marketing, promotion, and pricing decisions. In general, as the number of contingencies grows, so does the importance of decentralized decision making. A single decision maker might do a credible job identifying a profit-maximizing price in a peaceful, unchanging market. However, with scores of ever-changing market segments, setting prices centrally becomes daunting, perhaps even hopeless.

Let's now consider the contrary point of view and ask, under what circumstances does efficiency favor centralized decisions? The answer is, when decisions are highly interdependent, that is, when managers must coordinate choices. For instance, management's optimal output decision depends simultaneously on demand and cost. Thus, we should not delegate the output decision to a production manager (or a marketing manager) alone. Each would have only part of the relevant information to determine Q*. Accordingly, the output decision should be in the hands of more centralized managers who use demand and cost information from both the production and marketing segments of the firm. However, once centralized management has determined output, it can delegate many of the other decisions, such as the exact details of the advertising campaign, promotions, and input decisions, to the appropriate functional areas.

A second argument for centralization arises in the face of significant principal- agent problems. Imagine that a regional manager has the best information to make a particular decision, but that the manager's interests conflict with the firm's objectives. Absent controls or incentives to bend the manager's interests to the firm's, it would be foolish to delegate this decision. Instead, upper-level management should make the decision, even if it has imperfect information. Table 14.2 summarizes factors bearing on the choice between centralization and decentralization.

COORDINATION THROUGH TEAMS

Management's choice between centralization and decentralization is not all or nothing. The growing use of teams represents a hybrid sharing of information and decision responsibility. Teams pool information and perspectives. A 2007 study documented that the percentage

Factors Favoring Centralization	Factors Favoring Decentralization
1. High degree of coordination required	1. Low degree of coordination required
2. Concentration of decision-relevant information	2. Dispersion of decision-relevant information
3. Significant principal-agent problems	3. Compatible interests and objectives

of large firms with 20 percent or more of their employees in teams grew from 37 percent in 1987 to 61 percent in 1999. The evidence shows that the use of teams continues to increase today. For instance, like many large conglomerates, DuPont is organized with independent subsidiaries, each responsible for its own performance. In 2002, DuPont decided that it needed to offer national security products that cut across its wide range

of businesses. The company set up committees containing executives from across its subsidiaries to develop strategies and products to meet the new national security needs. Accenture, IBM, and Google are examples of companies that have used teams with great success. (IBM team members share information via the "wiki" technology.) ICU Medical has probably taken the team concept as far as any company. Any group of employees may form a team to solve a problem or to take on a project. Successful teams are rewarded monetarily and many have made substantial contributions.

COORDINATION VIA TRANSFER PRICES

Large, multi-division firms must coordinate activities among their divisions, which often provide goods or services to each other. For instance, a firm's automotive division might receive finished engines from its parts unit. Transfer prices—the prices that selling divisions charge to buying divisions within the firm—help coordinate internal actions. (See the appendix to Unit 6 for a full discussion and analysis.) The key to maximizing the firm's total profit is to set each transfer price equal to the marginal cost of the good or service in question. For instance, the automotive division should pay a transfer price for engines that reflects marginal cost; therefore, these costs are recorded dollar for dollar in computing the auto division's profitability. (Of course, accounting for all marginal costs is also crucial for setting optimal prices and quantities for finished vehicles.) As we saw in the case of Airbus, setting appropriate transfer prices is not always easy. The supplying division often pushes for a higher transfer price simply to enhance its own measure of profit, while the receiving division wants a lower price for the same reason. Either overstating or understating the transfer price can lead to incorrect decisions, resulting in underproduction or overproduction of both the transfer good and the final product. In short, setting transfer prices according to marginal cost is essential for efficient coordination within the multi-division firm.

MOTIVATING WORKERS

Probably the most pervasive form of principal agent relationship is that of employer and worker. For the relationship to work successfully, the firm (the principal) must motivate the worker (the agent) to act in the employer's interests. Workers have knowledge and abilities advantageous to their companies, but they also have their own needs and desires that may differ from the firm's objectives. Workers may simply wish to labor less hard and to enjoy life more, thereby sacrificing potential profits of the company.

Consider the following simplified model of the employer-worker relationship. The worker controls the amount of effort he or she puts into the job. An increased level of effort raises the workers' output, thereby increasing the company's profit. However, increased effort generates disutility for the worker.

Tying compensation to effort is one way to induce higher effort levels. If both the employer and employee can observe effort, they can design an optimal contract. First, employer and worker should agree to an effort level that maximizes the net benefit from the employment relationship. This net benefit is just _ _ D, where _ denotes the firm's profit and D

denotes the worker's disutility. Suppose, for example, that a second-year associate at a small law firm works an average of 55 hours per week in return for a $55,000 annual salary. Suppose that the associate generates $80,000 in additional net revenue for the firm and that the associate experiences a personal disutility valued at $40,000. Here, the net profit from the employment relationship is 80,000 _ 40,000 _ $40,000. The firm's share of this is 80,000 _ 55,000 _ $25,000, and the associate's share is $15,000 (the difference between the actual pay and the least amount he would accept in compensation for the disutility of the job). Both sides know these facts and know that working shorter or longer hours would diminish net profits. For instance, working 70 hours a week might generate increased billings and raise profit to $90,000, but it would also imply a disutility of $60,000, reducing the net benefit to $30,000. Thus, employer and worker settle on the efficient (55-hour-perweek) work arrangement.

EVALUATING GROUP PERFORMANCE

Frequently, group performance is easier to measure than individual performance for many of the reasons noted earlier. Rewarding group performance may encourage cooperation among employees who can all share in the fruits of their collective achievement. However, rewarding group performance does introduce new uncertainties into the compensation of the employee. That is, an employee's compensation and promotion are now tied to the efforts of others. More important, rewarding group performance may discourage optimal effort (indeed, encourage shirking), especially when the firm cannot easily observe individual effort. For example, suppose that a team consists of five workers whose annual bonuses will depend on the measured success of the group. Suppose that the group cannot observe individual effort but that if all team members exert percent extra effort, group performance will rise by a like amount and each member will reap an additional $25,000 in bonus. Because each member reckons the disutility of extra effort at $10,000, agreeing to become a high performing team benefits all (the net advantage is $15,000 each). But there is a catch. (Have you spotted the prisoner's dilemma yet?) Each member's personal incentive is to "free ride" on the efforts of the others. By exerting extra effort, the member raises the average performance of the group by only one-fifth of 15 percent, or 3 percent. (Remember, there are five team members.) In turn, the marginal individual benefit to exerting extra effort is only one-fifth of $25,000, or $5,000. The benefit of the extra effort is not worth the worker's disutility. The upshot of all attempting a free ride is that no one exerts extra effort.

Smaller groups naturally have fewer free-riding problems than larger groups, since it is easier to monitor effort by group members. In addition, mutual trust that all will do their part is easier to achieve in small groups. Despite potential problems with group compensation, the percentage of large firms that base some compensation on group performance has grown, as has the use of teams.

BOUNDARIES OF THE FIRM

Traditionally, economic activities have been undertaken within the firm because they are cheaper to do internally than through an external

market. As IT has reduced the cost of external transactions, firms have transferred a portion of in-house activities to external suppliers and markets. For instance, since the late 1990s, General Motors—the paragon of the traditional vertically integrated firm—has discontinued or spun off many of its component-manufacturing operations. (The spin-off of its Delphi Automotive Systems created a separate firm with $28 billion in sales.) Today, e-procurement systems allow companies to transact cheaply and efficiently with hundreds of external suppliers.

INTERNAL INFORMATION SHARING

Research has found that the greatest benefit of IT systems lies in facilitating information sharing within the firm. The chief of a conglomerate might rightly lament, "If our firm only knew what our firm knows." In other words, even when a firm is rich in its "knowledge capital," it may suffer from information overkill, loss, or waste. By contrast, an effective IT system allows efficient information sharing. Efficient information sharing can benefit both centralized and decentralized decision making. As we saw in the DHL case earlier, the efficient transmission of cost information to local managers improved the decision making for decentralized local managers. Likewise, efficient transmission of information can facilitate the movement of information up to centralized decision makers. In addition, ideas can be shared both within functional areas and between functional areas. For example, a production manager can share innovation in one assembly line with the firm's other six geographically dispersed lines. The marketing department can electronically make available customer information to the production and distribution departments. Besides her traditional departmental responsibilities, a marketing manager might work online as a member of an interdisciplinary team charged with managing a new-product launch. In many respects, IT systems foster decentralized decision making within firms. Delegation allows the local manager (with superior information and experience) to make appropriate decisions. At the same time, effective IT systems make it easier for higher company executives to monitor the local manager's actions and performance and to coordinate those actions with others' (if coordination is necessary). Because IT systems make it easier to monitor and measure worker performance, they go hand in hand with greater reliance on incentive pay for managers. In addition, IT systems tend to displace human managers in handling routine, rule-based job functions. For instance, today it takes only a handful of managers to operate and monitor a large-scale, automated cement-making plant. Separation of Ownership and Control in the Modern Corporation

An important example of the principal-agent problem occurs in large publicly held corporations. Such corporations are owned by vast numbers of shareholders (principals) and managed by directors (agents). Shareholders elect the board of directors, who oversee corporate management. This organizational form confers significant benefits in the financing of the firm. In issuing shares, the corporation gets access to a vast supply of financial capital, funds that would be difficult or impossible to secure from a single owner or even from a limited number of partners. Broad-based equity markets allow investors to diversify

across many firms and business sectors with the added protection of limited liability. (Limited liability means that the shareholders risk losing their investment, but no more than that. Creditors of the corporation cannot pursue the personal assets of shareholders.) However, shared ownership in the modern corporation does not imply shared control. In modern public corporations, shareholders do not have the right to manage. Setting day-to-day management decisions according to shareholder votes, besides being extraordinarily costly and impractical, would surely generate poor decisions. Rather, the organization vests decision-making responsibilities and control in a cadre of professional managers acting on behalf of shareholders. The problem is that shareholders have little practical control over the selection of top management or how top management performs once in place. Two roadblocks prevent shareholders from wielding voting power over the board and top management. First, management controls the voting and proxy process. (A shareholder uses a proxy to direct management in how to vote the shareholder's shares.) Top executives of U.S. corporations can use corporate funds to solicit proxies. By contrast, insurgent shareholders (those seeking to change management) receive compensation for their proxy solicitation only if they are successful in the proxy battle. The second obstacle stems from the difficulties of collective shareholder actions. Large institutional investors excepted, the typical shareholder owns a very small fraction of the outstanding voting shares of a large corporation. This shareholder recognizes that his or her vote will have a negligible effect on the outcome of any voting contest. Consequently, few shareholders will take the considerable time, effort, and cost of understanding the competing solicitations in a proxy battle. (This phenomenon is sometimes called rational apathy.) Most shareholders cast their votes for current management. Therefore, the chance is small that a challenge, no matter how meritorious, will succeed.

LIMITING THE POWER OF TOP MANAGEMENT

Because shareholders possess limited control over the selection and performance of top management, one would expect significant principal-agent problems. Top-level managers have the necessary information and presumably the expertise to make optimal decisions. However, managers often pursue their own agendas and undertake plans that conflict with the interests of shareholders. For instance, executives might engage in "empire building," thereby, incurring unnecessary costs (inflated management salaries, executive jets, and the like). Alternatively, in pursuit of the prestige of being market-share leaders, executives might be prone to expand the firm's operations far past the point of profit maximization. As the classical economist Adam Smith so eloquently put it The directors of such companies . . . , being the managers rather of other people's money than of their own, it cannot well be expected that they should watch over it with the same anxious vigilance with which the partners in a private copartnery frequently watch over their own. . . . Negligence and profusion, therefore, must always prevail, more or less, in the management of the affairs of such a company. Of particular concern is the area of executive compensation. By 2005 the ratio of CEO

pay to average worker pay in the United States had risen to 475 to 1, from around 24 to 1 in 1965. Compare the 2005 ratio to Britain (22:1), France (15:1), Germany (12:1), and Japan (11:1). The high level of executive compensation in the United States has not resulted in better economic performance relative to that of other countries. More likely, it is the sign of significant principal-agent problems. Because of these concerns, a number of mechanisms have arisen to mitigate the principal-agent problems inherent in large corporations.

1. Disclosure Requirements. Justice Louis Brandeis famously said, "Sunlight is the best disinfectant." Federal and state securities acts have rigorous disclosure requirements encompassing quarterly and annual reporting and disclosures in conjunction with proxy solicitation and tender offers. Indeed, the philosophy behind the Securities Act is primarily one of disclosure. If investors know the facts, they will be able to make sound financial decisions. In 2006, in response to ballooning executive pay, the Securities and Exchange Commission mandated greater disclosure to shareholders of executive compensation.

2. External Enforcement of Managerial Duties. In the United States, there are two types of enforcement. The first is enforcement through private rights of action. State and federal law give shareholders the right to sue if directors violate duties of care or loyalty or if directors engage in fraud, deception, or insider trading. Private attorneys have ample incentives through attorneys' fees to prosecute such cases on behalf of shareholders. The second mechanism involves direct government enforcement authorized by state and federal laws. For example, the government may sue or prosecute executives for deception, insider trading, or fraud.

FINANCIAL INCENTIVES

As noted in the previous section, incentive contracts can mitigate principal-agent problems. The same reasoning applies to a company's top management. By crafting pay-for-performance compensation plans, the organization can give managers greater incentives to maximize share value.26 This mechanism serves to reduce the

REVIEW QUESTION

1. What is asymmetric information?
2. Describe decentralization.
3. What is coordination through teams?
4. Describe coordination via transfer prices.
5. What is the process of motivating workers ?
6. What is the process of evaluating group performance?

FURTHER READINGS

1. Managerial Economics - Arun Kumar, Rachana Sharma
2. Managerial Economics - Thomas J. Webster
3. Managerial Economics - Petersen / Jain
4. Managerial Economics - Yogesh Maheshwari
5. Managerial Economics - E. Narayanan Nadar, S. Vijayan

UNIT-14 BARGAINING AND NEGOTIATION

CONTENTS

INTRODUCTION

Negotiation and bargaining are important features of many economic settings. Examples include negotiating the terms of a sales transaction, management labor bargaining, and settling a dispute out of court, to name just a few. Generally speaking, these are situations in which both parties stand to benefit from a cooperative agreement. Nonetheless, a significant degree of conflict remains because each side seeks to secure an agreement at terms most favourable to itself.

Many economic transactions are completed by means of bargaining under bilateral monopoly, that is, in settings in which a single seller faces a single buyer. In contrast to organized markets, in which competition among large numbers of buyers and sellers determines price and quantity, in bargaining settings the competition is one on one. Although the analysis of market competition obviously deserves attention (see Units 7, 8, and 9), it is worth remembering that there are other important means of resource allocation. Our objectives in this Unit are twofold. In the first two sections, we analyse the economic forces underlying the bargaining setting: What economic factors create the opportunity for mutually beneficial agreements? What form do economically efficient bargains take? Next, we examine bargaining strategy from the perspective of decision making under uncertainty: What bargaining strategy maximizes management's expected profit from the transaction? What are the risks of such a strategy? Finally, we apply the principles of

negotiation to the historic takeover dispute between Texaco and Pennzoil.

THE ECONOMIC SOURCES OF BENEFICIAL AGREEMENTS

It takes two to tango and three to form a ménage á trois. In other words, economic agents enter into transactions because the transactions are mutually beneficial. A well-crafted agreement is better for both parties than no agreement at all. Moreover, some agreements are better (for both parties) than others. Given this observation, it is natural to explore the economic factors that create the opportunities for mutually beneficial agreements. We begin our discussion by considering a typical negotiated transaction involving a buyer and a seller.

SELLING A WAREHOUSE

Two firms are locked in negotiations concerning the sale of a warehouse, the equipment therein, and a considerable inventory of industrial machinery. The main issue is price. The present owner is closing down its current operation in a move to redirect its resources into other businesses. The warehouse is in a valuable location for the would-be buyer, who also could make direct use of the equipment and machinery inventory. The buyer has examined the warehouse and contents and, after considerable figuring, has estimated its value for the transaction at $600,000; that is, the potential buyer is indifferent to the options of paying $600,000 to complete the purchase or forgoing the transaction altogether. The seller sets its value for the transaction at $520,000; this is the net amount the firm estimates it would obtain, on average, from selling the warehouse and contents via a broker or at auction. The buyer and seller values are referred to as reservation prices or walk-away prices. Given the values held by buyer and seller, it is evident that a mutually beneficial agreement is possible. In particular, both parties would prefer an agreement at a price between $520,000 and $600,000 to the alternative of no agreement at all. For convenience, we denote the sale price by P. The seller's profit from such a transaction is P _ $520,000, whereas the buyer's gain is $600,000 _ P. If there is no agreement on a price (and, therefore, no sale), each party earns zero profit. Clearly, any price such that $520,000 _ P _ $600,000 affords positive profits for both parties. This price range between the buyer and seller walk-away prices is referred to as the zone of agreement. Observe that the total gain (the sum of buyer and seller profit) from such a transaction is .

The total gain (or trading gain) is measured by the difference between the buyer and seller values, that is, the size of the zone of agreement. 15.1 presents two views of the buyer-seller transaction. Part (a) shows the zone of agreement and possible negotiated prices within it. A price of $540,000 is shown at point A. At this price, the buyer claims $60,000 in profit and the seller claims $20,000. Obviously, at higher negotiated prices, the seller's profit increases and the buyer's profit falls dollar for dollar. Part (b) displays this profit trade-off explicitly. The parties' profits from transactions at various prices are graphed on the axes. The profits from a $540,000 price appear at point A. Prices of $560,000 and

$580,000 (and the corresponding profits) are listed at points B and C, respectively. The downward-sloping line shows the profit implications for all possible prices within the zone of agreement. This is commonly called the payoff frontier. If the parties fail to reach an agreement, they obtain zero profits, as marked by point 0 at the origin of the graph. 15.1 reemphasize a simple but important point about the gains from a negotiated agreement. An agreement at any price between $520,000 and $600,000 is better for both parties than no agreement. The "no agreement" outcome is said to be inefficient because there exists one or more alternative outcomes that are better for both parties. We say that an outcome is efficient if no other outcome exists that is better for both parties.1 By this definition, all of the out- (600,000 _ P) _ (P _ 520,000) _ 600,000 _ 520,000 _ $80,000

More accurately, an agreement is efficient if there is no other agreement that makes one party better off without making the other worse off.

(b) Player Profits occurs is a matter of competition: The buyer prefers a low price, the seller a high price. In the negotiation literature this situation is called a distributive bargainbecause the parties can be thought of as bargaining (via price) over the distribution of the total profit (in this case, $80,000) available from the transaction. The actual price they negotiate depends in part on the bargaining abilities of the parties and on notions of equity and fairness. For instance, a final price in the vicinity of $560,000 (implying $40,000 in profit for each side) might be negotiated by equally matched bargainers who are in agreement that the total bargaining profit should be divided equitably. For the moment, however, our analysis has identified the zone of agreement without offering a prediction of which price within this zone will be the agreement terms.

Two additional points can be drawn from the example. First, the source of the trading gains is the difference in the parties' values. Because the seller's value for the warehouse and items is less than the buyer's value, completion of the transaction creates a trading gain that both sides share. In contrast, if the agent values were reversed (i.e., the seller's value was $600,000 and the buyer's value $520,000), no mutually beneficial transaction would be possible. Second, the values or reservation prices of the parties are influenced by the alternative transactions available to them. In the present circumstances, for instance, the buyer estimates the monetary value for the warehouse at $600,000. Clearly, if the buyer learned of the availability of another warehouse at a comparable location at an unexpectedly low price, its walk-away price for the current transaction would fall markedly. Similarly, if the buyer revised downward its estimate of the potential profit from the warehouse operation (because of adverse economic conditions in general), its walk-away value also would fall. Of course, the importance of outside opportunities pertains equally to the seller. In short, the alternative transactions available to the parties directly or indirectly set the respective walk-away prices between which negotiated agreements can occur.

PROBABILITY ASSESSMENTS

Even if two parties have identical preferences, they may assess different values for a transaction due to different probability assessments and forecasts. For instance, an agreement may be supported by each side's optimistic belief that the transaction is substantially better than no agreement at all. As Mark Twain said, "It is differences of opinion that make horse races." Many transactions involve an element of a bet: Each side believes it has a better assessment of the transaction's value than the other and will gain (possibly) at the other's expense. Of course, differences in probability assessments also can work against negotiated agreements. The following application makes the point.

SETTLEMENT NEGOTIATIONS REVISITED

Let's return to the patent dispute, but now suppose the firms hold different, conflicting assessments about the litigation value of the case. The small firm believes there is a .6 chance that its side will win the case (i.e., there will be a finding of patent infringement). The large firm assesses a .6 chance that it will win the case (i.e., no infringement will be found). Both sides estimate an expected damage award of $2 million for an infringement finding and no damages otherwise. Therefore, the parties' expected values are: vS _ (.6)($2.0) _ $1.2 million and vL _ (.4)(2.0) _ $.8 million. Accounting for the parties' legal costs (as in constraint 15.1), the least the small firm will accept out of court is $1 million, whereas the most the large firm will offer is $960,000. Thus, there is no zone of agreement. In general, a negotiated settlement is possible if and only if there is some price, P, such that constraint 15.1 is satisfied. An equivalent constraint is which is derived by rearranging constraint 15.1. A mutually beneficial settlement is possible if and only if the difference between the parties' litigation expectations is smaller than the combined court costs.

RISK AVERSION

Recall from the discussion in Unit 12 that a risk-averse agent assesses the value for an uncertain outcome to be significantly lower than its expected value (EV). This value is termed the outcome's certainty equivalent (CE). In algebraic terms, CE _ EV. The greater the agent's risk aversion and/or the riskiness of the outcome, the greater the gap between the certainty equivalent and expected value.

The presence of risk aversion motivates transactions that minimize and/or distribute risks among the parties. For instance, consider the patent dispute once again. We saw that, when each side assessed its winning chances at 60 percent, the parties' expected payoffs (court costs included) were $1 million and _$960,000; thus, no settlement was possible. However, because the litigation outcome is highly uncertain, we can expect each risk-averse disputant to value going to court at a CE value considerably below its EV. For example, suppose the small firm judges its CE value for going to court at $800,000 (including court costs), and the large firm sets its CE at _$1.1 million. Now there is a $300,000-wide zone of agreement in the settlement negotiations. The presence of risk aversion makes a certain out-of-court settlement more attractive than a risky outcome in court (even though each side is

optimistic about the outcome at trial). As a general principle, transactions should be designed so that risks are assumed by the party best able to bear them. Consider the wildcatter in Unit 12 who holds an option to drill for oil on a geological site. Suppose the wildcatter estimates the expected profit of the site to be $140,000 but, being risk averse, assesses the CE value of the site to be considerably less than this—say, $100,000. Should the wildcatter explore the site or sell the option to a giant exploration company that drills scores of wells in all parts of the world? Suppose the large drilling company is risk neutral. If its geologists agree with the wildcatter's probabilistic assessments, the company's value for the site is $140,000. Consequently, the option can be sold at a mutually beneficial price between $100,000 and $140,000. The option should be transferred to the risk-neutral party because that party values the site more highly. A classic case of a transaction designed for optimal risk bearing is the cost plus contract used in high-risk procurements. The risks concerning performance, cost, and timetable of delivery in defense procurement—for instance, in the development of a new weapons system or aircraft—are enormous. As a result, the usual fixed-price contract, in which the defense contractor is paid a fixed price and bears all production risk, is impractical (that is, the firm would set an extremely high fixed price—add a substantial risk premium—to compensate for possible cost overruns). Given its vast financial wealth, the federal government arguably can be characterized as risk neutral. The government, rather than the firm, should bear the contract risk. Under a cost-plus contract, the government reimburses the firm for all allowable costs and pays it a fixed profit amount in addition. The large variability in cost is borne by the government buyer, whereas the contractor's profit is guaranteed. The government benefits by paying the firm a much lower profit fee than would be required if the firm were the risk bearer. When both parties are risk averse, the optimal response to uncertainty is risk sharing. Returning to the oil example, suppose a second drilling firm is identical to the first; that is, it is equally risk averse and holds the same probability assessments. Then the site has a CE value of $100,000 to either party. Because here is no difference in value, there is no possibility of mutual benefit from an outright sale. But consider what happens if the two companies form a partnership to share equally (i.e., 50-50) all profits and losses from drilling. The expected value of each side's 50 percent profit share is, of course, $70,000. What is each side's CE for its share? Because each outfit now is exposed to considerably smaller. One disadvantage of the cost-plus contract is that it offers the firm very little incentive to keep costs down.

CONTINGENT PRICING IN AN ACQUISITION

Firm A is negotiating to buy a division of firm T. The difficulty is that the value of the division depends on whether it wins the bidding for a major contract from the government. If it wins, the division will be worth $20 million under current management and$22 million if acquired by firm A. If it loses, it will be worth $10 million under current management and $12 million if acquired by firm A. In either case, the division is worth more to firm A than to firm T, due to synergies with firm A's other operations. Firm T judges a .7 probability that the division will

win the contract, but firm A judges this probability to be only .4. Is a mutually beneficial agreement possible?

To answer this question, first consider a straight cash buyout. Firm T values the division at (.7)(20) _ (.3)(10) _ $17 million. The price must be at least this high to be acceptable. Firm A computes the expected value at (.4)(22) _ (.6)(12) _ $16 million, so it will pay no more than this. Consequently, a cash buyout is impossible. Both sides agree that the division will be worth more under firm A than under firm T (regardless of the contract outcome). But the parties' different, conflicting probability assessments make a straight cash purchase impossible.

However, the acquisition can be consummated if a contingent-pricing clause is included. Suppose the parties agree that the purchase price will be $21 million if the government contract is won and $11 million if it is not. Clearly these price terms provide each side a $1 million profit regardless of the government contract outcome. Contingent pricing neatly overcomes the obstacle posed by conflicting probability beliefs.

The use of contingent contracts is a common response to risk and uncertainty in purchase and sale arrangements. Warranties and guarantees are obvious examples. Here the terms of the agreement are adjusted in light of future events. Another response to uncertainty is the use of incentive contracts, which call for both buyer and seller to share the burden of cost overruns. Acquisition of an enterprise at a purchase price that depends on the firm's future earnings is still another example. Corporate acquisitions paid for with securities of the acquiring firm embody an element of contingent pricing. If the acquisition is truly valuable, the securities of the merged company will appreciate.

MULTIPLE-ISSUE NEGOTIATIONS

Thus far, we have considered single-issue agreements in which price is the only object of the negotiation. Here an agreement within a range of prices is mutually preferred to no agreement at all. The negotiation setting becomes more complicated when the terms of an agreement involve multiple issues, such as performance specifications, service requirements, or product attributes, as well as price. When multiple issues are at stake, the parties cannot be satisfied in simply finding an agreement; rather, the goal is to uncover an optimal agreement— one that, roughly speaking, is best for both parties. Even if the parties have conflicting interests on each of many separate issues, diligent negotiations can arrive at a well-crafted agreement that is better for both sides than alternative agreements. The simplest of examples suffices to make the point. Consider two members of a legislative committee whose interests are directly opposed on each of two issues. Ms. A strongly favors issue 1 and weakly opposes issue 2. Mr. B strongly favors issue 2 and weakly opposes issue 1. Can these members fashion a mutually beneficial voting agreement? The answer is yes. They should agree to "swap votes" so that both vote affirmatively on each issue. By gaining a vote on the issue that is more important to him or her, each member is better off after the swap (even though the member votes against his or her strict self-interest on the unimportant issue). This example illustrates a principle that is applicable to bargaining in general:

In multiple-issue negotiations, as long as there are differences in the value (importance) parties place on issues, there will be opportunities for mutually beneficial agreements of a quid pro quo nature.

In multiple-issue bargaining involving monetary transfers, the key to the attainment of efficiency is to structure agreements to maximize the total value the parties derive from the transaction. The logic of this result is quite simple. The transacting parties should form an agreement that maximizes the size of the profit "pie" to be split. Then negotiation of an overall price for the transaction has the effect of dividing the pie between the parties. Any such division of the maximal total value is efficient; one side cannot gain without the other side losing. In turn, any division of a less-than-maximal total value is necessarily inefficient. An appropriately priced maximal-value agreement delivers higher profits for both sides. We offer a concrete example to illustrate this result.

A COMPLEX PROCUREMENT

The Department of Defense (DOD) is in the process of negotiating a procurement contract for aircraft engines with an aeronautics firm. The contract will specify the number of engines to be delivered, the time of delivery, and the total price to be paid by DOD to the contractor. The firm has assessed its total cost of supplying various quantities of engines by different deadlines. For its part, DOD has assessed monetary values (its maximum willingness to pay) for different contracted deliveries. Table 15.1 lists the parties' costs and values. Suppose DOD and the firm are considering a contract for 40 engines in four years at a price of $39 million. Is this contract mutually beneficial? Could both parties do better under a different contract at the right price? Of the nine possible combinations of order sizes and delivery dates, which should the parties adopt?

From Table 15.1, we find the parties' profits under the 40-engine, four year contract ($39 million price) as follows: The firm's profit is 39 _ 36 _ $3 million; DOD's profit is 42 _ 39 _ $3 million. Clearly, this is a mutually beneficial agreement. However, it is evident from the table that the parties can improve on these contract terms. The value-maximizing contract calls for 80 engines to be delivered in three years. This contract offers a total profit of 85 _ 70 _ $15 million. (This is just the difference between DOD's value and the firm's cost.) At a $77.5 million price, each side earns a $7.5 million profit— some two-and-one-half times the profit under the four-year, 40-engine agreement. The three-year, 80-engine contract is efficient. All other contracts offer lower total profits and, therefore, are inefficient.

In negotiating a contract, firm A and firm B are considering three options. Firm A can supply firm B with a 97% pure compound, a 98% pure compound, or a 99% pure compound. Raising purity by 1% increases firm A's cost by $50,000. Firm B's potential profits are $200,000, 280,000, and $320,000 for the 97%, 98%, and 99% compounds. Which of the three options constitutes an efficient agreement (i.e., that maximizes the parties' total value or "pie")? Why?

A QUANTITY-PRICE CONTRACT

A buyer and seller are negotiating the terms of a delivery contract specifying price and output quantity (Q). The buyer's total value from

purchasing Q units is B _ 3Q _ Q2/20. The seller's cost of producing Q units is C _ Q2/40. The parties seek an agreement as to the quantity, Q, and the total payment from buyer to seller (call this R). What order quantity is part of an efficient agreement?

A direct way to characterize an efficient agreement is to find the value maximizing order quantity. The sum of buyer and seller profits is Total net benefit (B _ C) is maximized by setting marginal benefit equal to marginal cost: MB _ dB/dQ _ 3 _ Q/10 and MC _ dC/dQ _ Q/20. Setting these equal to each other gives Q _ 20. At this quantity, the buyer's benefit is B _ 3(20) _ 202/20 _ 40, and the seller's cost is C _ 202/40 _ 10. The relevant negotiation region for the payment, R, is the range between 10 and 40, and the maximum total profit is 40 _ 10 _ 30. (This assumes each party faces a zero profit from a disagreement; i.e., each has no other profitable alternative.) A graphical analysis provides additional insight into the meaning of efficiency when continuous variables are the object of negotiation. In 15.2, the axes list the variables, Q and R. Thus, any point on the graph represents possible terms of an agreement. The next step is to show the profit implications of any agreement. This is done by means of profit contours, the series of curves in the .6 The black curves show the seller's profit contours; the colored curves are the buyer's. For instance, the lowest seller contour (marked _S _ 0) shows all combinations of Q and R that provide exactly a zero profit. This is identical to the firm's cost curve: R _ C _ Q2/40. The curve is upward sloping; to maintain a zero profit, the firm must receive a higher R for producing a larger Q. The next highest contour (_S _ 7, only part of which is shown) shows Q and R combinations yielding a profit of 7. In general, higher profit contours are simply vertical displacements of lower ones. The seller's (B _ R) _ (R _ C) _ B _ C. In a great many economic settings, a slightly different terminology is used. 15.2 often is called an Edgeworth box, and the contours are called indifference curves. For instance, we examined an individual's indifference curves in the appendix to Unit 3. When the individual gains from an increase in either variable, the indifference curves will be downward sloping. (To leave the individual indifferent, a reduction in one variable must be compensated by an appropriate increase in the other.) profit increases as one moves north or west in the diagram, that is, as R increases (for fixed Q) or Q falls (for fixed R). The interpretation of the buyer's contours (the colored curves) is analogous, but the orientation is reversed: the buyer profits from lower R and/or higher Q, that is, from south and east movements in 15.2. In particular, note that the zero-profit contour is uppermost in the and that the buyer's profit increases with moves to lower contours.7 How can we use these profit contours to identify efficient agreements? The answer is provided by the following important result: An agreement is efficient if, and only if, it lies on buyer and seller profit contours that are tangent to each other.

NEGOTIATION STRATEGY

Negotiations inevitably produce tension between the forces of competition and cooperation. To reach a mutually beneficial agreement, both sides must cooperate. More than that, they must strive to uncover better agreements. Yet each side's ultimate objective is to secure the

most favorable agreement for itself. Of course, along the payoff frontier securing better terms for oneself implies less favorable terms for the other side. Thus far, our discussion has focused on identifying efficient agreements, that is, outlining the best the parties can do together. However, for a variety of reasons, bargaining as actually practiced often falls far short of optimal outcomes. In his seminal work on bargaining, The Strategy of Conflict, Thomas Schelling puts the problem this way: Most bargaining situations ultimately involve some range of possible outcomes within which each party would rather make a concession than fail to reach agreement at all. In such a situation any potential outcome is one from which at least one of the parties, and probably both, would have been willing to retreat for the sake of agreement, and very often the other side knows it. Any potential outcome is therefore one that either party could have improved by insisting; yet he may have no basis for insisting, since the other knows or suspects that he would rather concede than do without an agreement. Each party's strategy is guided mainly by what he expects the other side to insist on; yet each knows that the other is guided by reciprocal thoughts. The final outcome must be a point from which neither expects the other to retreat. To put this another way, any set of terms falling inside the zone of agreement can be supported as an equilibrium outcome. As an example, consider two parties bargaining over the division of the total profit from a mutually beneficial transaction. Bargaining takes place in the simplest possible way: Each side makes a single offer, naming his or her share of the total profit. If the offers are compatible (i.e., they add up to less than 100 percent of the total profit), there is an agreement (each party getting his or her offer); otherwise, there is no agreement. Here any pair of offers summing to exactly 100 percent constitutes an equilibrium. For instance, offers of 50 percent each are in equilibrium. Neither side can profit by (1) demanding more, because this leads to a disagreement and zero profit, or (2) demanding less, because this directly lowers his or her profit. In turn, the offers 80 percent and 20 percent (or any other pair of compatible offers, no matter how inequitable) are also in equilibrium. The cold truth is that, against an opponent whose nonnegotiable demand is for 80 percent of the profit, the best one can do is settle for the remaining 20 percent. To sum up, any division of the profit (equitable or inequitable) is an equilibrium outcome.

Via the dynamic process called bargaining, parties will arrive at some final outcome. But the multitude of equilibrium outcomes makes it difficult to predict which one. Clearly the final outcome depends significantly on the bargainers' expectations—expectations that are modified via the exchange of offers and counteroffers during the negotiations. In some sense, bargaining ceases when expectations converge, at a point where neither side can expect the other to concede further. Then either an agreement is signed or, if the sides stubbornly hold to conflicting expectations, a disagreement results.

Perfect Information

If both sides have perfect information—that is, there is no uncertainty about the economic facts of the negotiation—profit-maximizing bargainers always 8T. C. Schelling, The Strategy of Conflict

(Cambridge, MA: Harvard University Press, 1990). should reach an efficient agreement. The reason is simple. To settle for an inefficient agreement is to leave money on the table. This cannot be profit maximizing; there exist alternative terms providing greater profit for both parties. As we saw earlier, if the disputants in a conflict are sure of the disposition of the case if it goes to court, they should settle the case in the first place with both benefiting from saving the collective costs of going to court. At the same time, we should emphasize that what is true in theory does not always hold in practice. Even under perfect information, identifying and implementing efficient agreements is far from easy.

Imperfect Information

Will each bargainer typically have perfect information about the benefits and costs (both to itself and its bargaining partner) of potential agreements? A more realistic description of the bargaining setting posits imperfect information on the part of the bargainers. Typically each side has only limited information about its own values for potential agreements and, at best, will have only probabilistic information about the other side's values. Under imperfect information, issues of bargaining strategy become increasingly important. In a simple price negotiation, for instance, neither side knows for certain how far it can push the other before an agreement becomes impossible. Indeed, neither can be certain whether there is a zone of agreement. The negotiation process itself conveys information about possible acceptable agreements, but this information cannot be taken at face value. In everyday bargaining, the parties typically start with exaggerated and incompatible demands. It would be foolish for one side to concede immediately to the other's opening offer. Similarly, it would be unwise for one side to "lay its cards on the table" and reveal its true value for the transaction at the outset. In short, bargaining strategy calls for a significant element of bluff.

The theory of negotiation under uncertainty yields an important result: In bargaining settings under imperfect information, optimal bargaining behaviour is incompatible with the attainment of efficient agreements all of the time. Imperfect information presents a barrier to the attainment of efficient agreements both during and after the actual negotiations. As we shall see, it generally is in the self-interest of each side to keep its values private—indeed, to misrepresent its values during the negotiations for the purpose of assuming a "tough" bargaining stance. The result is a predictable number of missed and/or inefficient agreements. The presence of uncertainty after an agreement is signed also poses problems. For instance, if agreements are difficult to monitor or enforce, there may be insufficient incentives for one or both parties to fulfill the terms of the agreement. The following example shows clearly how optimal bargaining behavior can result in a failure to attain certain beneficial agreements.

A TENDER OFFER

Firm A (the acquirer) is about to make a first-and-final price offer for the outright purchase of family-owned firm T (the target). Firm A is confident the target will be worth $1.6 million under A's own

management. It has only a vague idea of firm T's reservation price, that is, the minimum price current management will accept. Its best guess is that this value (denoted by v) is uniformly distributed between $1 million and $2 million; that is, all possible values in this range are equally likely. What is the firm's best offer? How often will a sale be concluded?

Clearly the acquirer can confine its attention to offers in the $1 million to $1.6 million range. Firm A faces an obvious trade-off between the probability and profitability of agreements. The higher its offer, the greater the chance of acceptance, but the lower the transaction profit. The firm's expected profit from offer P is Here, we have used the fact that Pr(P is accepted) _ P _ 1. For instance, as predicted by this expression, the offer, P _ $1.5 million, is accepted half the time (by a target with a value anywhere between $1 million and $1.5 million). The higher offer, P _ $1.8 million, is accepted with probability .8, and so on. To maximize expected profit, we set Thus, the optimal offer is P* _ $1.3 million. The probability that this price will be accepted is .3, implying that the acquirer's maximum expected profit is $90,000. The point to underscore is this: The acquirer maximizes its expected profit by taking a calculated risk; it shades its offer well below its true value, even though this tactic poses the risk of missing possible agreements (whenever the target's reservation price is between $1.3 million and $1.6 million). The lesson of this example carries over to the case of multiple offers and counteroffers. In equilibrium, a self-interested bargainer always should hold out for terms that are strictly better than its true reservation price, thereby incurring the risk that some possible agreements are missed. Put another way, suppose one side always is willing to concede up to its true value, if necessary, to reach an agreement. Clearly, the other side could take advantage of this purely cooperative behavior by "waiting the player out"—agreeing to terms only after the player has made full concessions. To protect itself against this "waiting" strategy, a player must be willing to risk disagreement. As movie producer Sam Goldwyn once said, "The most important thing in acting is honesty. Once you've learned to fake that, you've got it made." To a degree, the same can be said of bargaining: Under imperfect information, a certain amount of dissembling, playing one's cards close to the vest, is essential. Otherwise, one is prone to the danger of being read like an open book by an opponent.

Repetition and Reputation

Thus far, we have focused on a one-time negotiation between a pair of interested parties. As a natural consequence, the parties' bargaining behavior has been motivated solely by the immediate profit available from an agreement. Now let's consider the effect if one or both parties are expected to face different bargaining situations repeatedly. For instance, labor contracts typically are no longer than three years. Thus, even when the current contract is signed and sealed, labor and management are well aware they will be negotiating a new contract in two or three years' time. Alternatively, one side may find itself repeatedly negotiating with scores of different parties over time. As an example, representatives of insurance companies negotiate hundreds of tort and liability claims each year.

Repeated negotiation (with the same or different parties) introduces the key strategic element of reputation; that is, the firm recognizes that its behaviour in the current set of negotiations can influence the expectations of its future bargaining partners. In a one-time bargaining setting, in contrast, the firm's actions are motivated solely by immediate profit; issues of reputation do not enter. One important effect of reputation formation in repeated negotiations is to limit the scope of purely opportunistic behavior. To illustrate, consider current contract negotiations between two firms, A and B. Due to many bargaining factors in its favor, A is confident it can negotiate a contract giving it 90 percent of the total profit from an agreement. If it expects never to bargain with B again, A surely will push for these favorable terms. But what if B and A are likely to bargain with each other over many subsequent contracts?

Negotiating too good a contract poses the risk of souring the entire bargaining relationship. (Perhaps B would spurn A and seek out a new bargaining partner in the future.) Accordingly, A may rationally choose not to take full advantage of its short-term bargaining power.

Reputation effects also suggest that B, the weaker bargaining party, may be unwilling to concede the lion's share of the short-term gain to A. In a one shot bargain, accepting 10 percent of something is better than nothing. But in repeated bargaining, B must be concerned about its reputation. Large concessions now may spur the other party to take a tougher bargaining stance in the future. Thus, B has an interest in establishing a reputation as a tough but fair bargainer. Sometimes this reputation effect means sacrificing or delaying short-term agreements. For instance, strikes frequently occur because one or both sides seek to establish their long-term reputations. Insurance companies typically take a tough stance toward settling claims of uncertain merit. Viewing the claim by itself, the company might find it cheaper to settle than to go to court. Nonetheless, on reputation grounds, it pays to fight to deter questionable claims in the future.

Finally, the repeated bargaining relationship has a disciplining role—a role we already noted in Unit 10 in our discussion of the repeated prisoner's dilemma. Recall that, in the one-shot prisoner's dilemma, the dominant-strategy equilibrium calls for noncooperation. In contrast, in the infinitely repeated prisoner's dilemma, continual cooperation is an equilibrium. The key to this equilibrium is one side's credible threat to punish the other's noncooperation with a retaliatory response. In short, bargaining partners that are "married" to each other have obvious incentives to maintain a cooperative relationship.

Business Behavior: Failed Agreements

In actual business practice, negotiation behavior predominantly follows economic predictions. For instance, when both bargainers have complete information about the mutual benefits to be had from a successful transaction or deal, an agreement should soon follow. Contrary to the notion of the litigious American legal system, most disputes (some researchers estimate more than 90 percent) end in amicable settlement agreements rather than costly court proceedings. Deal makers routinely trade off multiple issues and include contingent clauses as needed in

order to increase the total value of an agreement, which the parties can then share. Nonetheless, there are instances when bargaining behavior and outcomes diverge from the textbook predictions advanced by economic principles. Highprofile disputes occur even when there is strong evidence of mutual benefit from a timely agreement. Though less frequent today, costly strikes—the lengthy screenwriters strike in Hollywood and the National Hockey League impasse causing the 2004–2005 season to be canceled—persist. Most strikes are ultimately settled at terms that could have been concluded much earlier, without incurring the attendant economic costs to both sides. Frequently, top management of a target company rebuffs a merger or takeover advance, even when it would deliver a large price premium to shareholders. Chief executive officers (the venerable Jack Welch included) are tangled in costly and public divorce disputes. The death of a business mogul triggers an ugly dispute about how his or her inheritance and control of the family business should be divided among layers of the family tree.

Recall our earlier point that disputants who hold the same information about the case under contention should always find their way to an efficient, mutually beneficial agreement (assuming one exists). But several factors impede agreements in practice. First, research by psychologists has documented a key impediment to agreements: self-serving bias. For instance, consider a bargaining experiment in which participants are assigned the roles of plaintiff and defendant in a legal case and are given exactly the same facts and information. The economic prediction is that the parties (sharing the same valuation of the case) should always agree to a settlement to avoid the legal costs of going to court. The results are quite the contrary. Invariably, the plaintiff sees a much greater court award than does the defendant—on exactly the same evidence. One's prediction is biased (consciously or unconsciously) by one's self-interest. Therefore, disputants in these experiments frequently litigate and incur the associated court costs. A second source of missed agreements occurs when there are multiple mutually beneficial bargaining equilibria. Recall the conflict over the standard for high-definition DVDs described in Unit 10, Table 10.4. For years, each side adamantly held to its preferred incompatible format, severely impeding adoption of either new technology. Here, the problem was a failure to agree on either equilibrium. Third, notions of fairness can aid or impede agreements. On the one hand, the fairness of a transparent 50-50 split can offer an obvious point of agreement. On the other, there might be many possible candidates for a fair agreement, among which the bargainers disagree. Suppose that agreement A is one such agreement candidate that both sides prefer to their disagreement outcomes. Nonetheless, it is not unusual for bargaining to end in disagreement, simply because one side finds agreement A unacceptable on grounds of fairness. Finally, as we have noted, when bargainers hold imperfect information, selfinterested negotiation behavior leads to missed agreements, at least some of the time. Thus, some frequency of disagreements should not be surprising in settings where each side has only partial information about potential agreements. Indeed, other means of dispute resolution such as mediation and arbitration are sometimes invoked to facilitate the prospect of

reaching an accord. To sum up, bargaining is a valuable, but not perfect, means of reaching agreements.

REVIEW QUESTION

1. Describe the economic sources of beneficial agreements.
2. What is the process of selling a warehouse
3. What is probability assessments
4. What is risk aversion
5. What is multiple-issue negotiations
6. Describe negotiation strategy for
7. What are the difference between perfect information and imperfect information?

FURTHER READINGS

1. Managerial Economics - Arun Kumar, Rachana Sharma
2. Managerial Economics - Thomas J. Webster
3. Managerial Economics - Petersen / Jain
4. Managerial Economics - Yogesh Maheshwari
5. Managerial Economics - E. Narayanan Nadar, S. Vijaya

BIBLIOGRAPHY

6. Managerial Economics - Arun Kumar, Rachana Sharma
7. Managerial Economics - Thomas J. Webster
8. Managerial Economics - Petersen / Jain
9. Managerial Economics - Yogesh Maheshwari
10. Managerial Economics - E. Narayanan Nadar, S. Vijayan
11. Managerial Economics - Managerial Economics - Ivan Png
12. Managerial Economics - Mark Hirschey