READING

13

The Firm and Market Structures

by Richard Fritz, PhD, and Michele Gambera, PhD, CFA

Richard Fritz, PhD, is at the School of Economics at Georgia Institute of Technology (USA). Michele Gambera, PhD, CFA, is at UBS Asset Management (Americas), Inc. (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** describe characteristics of perfect competition, monopolistic competition, oligopoly, and pure monopoly;
☐	**b.** explain relationships between price, marginal revenue, marginal cost, economic profit, and the elasticity of demand under each market structure;
☐	**c.** describe a firm's supply function under each market structure;
☐	**d.** describe and determine the optimal price and output for firms under each market structure;
☐	**e.** explain factors affecting long-run equilibrium under each market structure;
☐	**f.** describe pricing strategy under each market structure;
☐	**g.** describe the use and limitations of concentration measures in identifying market structure;
☐	**h.** identify the type of market structure within which a firm operates.

INTRODUCTION

1

The purpose of this reading is to build an understanding of the importance of market structure. As different market structures result in different sets of choices facing a firm's decision makers, an understanding of market structure is a powerful tool in analyzing issues such as a firm's pricing of its products and, more broadly, its potential to increase profitability. In the long run, a firm's profitability will be determined by the forces associated with the market structure within which it operates. In a highly competitive market, long-run profits will be driven down by the forces of competition. In less competitive markets, large profits are possible even in the long run; in

the short run, any outcome is possible. Therefore, understanding the forces behind the market structure will aid the financial analyst in determining firms' short- and long-term prospects.

Section 2 introduces the analysis of market structures. The section addresses questions such as: What determines the degree of competition associated with each market structure? Given the degree of competition associated with each market structure, what decisions are left to the management team developing corporate strategy? How does a chosen pricing and output strategy evolve into specific decisions that affect the profitability of the firm? The answers to these questions are related to the forces of the market structure within which the firm operates.

Sections 3, 4, 5, and 6 analyze demand, supply, optimal price and output, and factors affecting long-run equilibrium for perfect competition, monopolistic competition, oligopoly, and pure monopoly, respectively.

Section 7 reviews techniques for identifying the various forms of market structure. For example, there are accepted measures of market concentration that are used by regulators of financial institutions to judge whether or not a planned merger or acquisition will harm the competitive nature of regional banking markets. Financial analysts should be able to identify the type of market structure a firm is operating within. Each different structure implies a different long-run sustainability of profits. A summary and practice problems conclude the reading.

2 ANALYSIS OF MARKET STRUCTURES

Traditionally, economists classify a market into one of four structures: perfect competition, monopolistic competition, oligopoly, and monopoly. Section 2.1 explains that four-way classification in more detail. Section 2.2 completes the introduction by providing and explaining the major points to evaluate in determining the structure to which a market belongs.

2.1 Economists' Four Types of Structure

Economists define a market as a group of buyers and sellers that are aware of each other and can agree on a price for the exchange of goods and services. While the internet has extended a number of markets worldwide, certain markets are limited by geographic boundaries. For example, the internet search engine Google operates in a worldwide market. In contrast, the market for premixed cement is limited to the area within which a truck can deliver the mushy mix from the plant to a construction site before the compound becomes useless. Thomas L. Friedman's international best seller *The World Is Flat*[1] challenges the concept of the geographic limitations of the market. If the service being provided by the seller can be digitized, its market expands worldwide. For example, a technician can scan your injury in a clinic in Switzerland. That radiographic image can be digitized and sent to a radiologist in India to be read. As a customer (i.e., patient), you may never know that part of the medical service provided to you was the result of a worldwide market.

Some markets are highly concentrated, with the majority of total sales coming from a small number of firms. For example, in the market for internet search, three firms controlled 98.9 percent of the US market (Google 63.5 percent, Microsoft 24 percent, and Oath (formerly Yahoo) 11.4 percent) as of January 2018.[2] Other markets are

1 Friedman (2006).

2 Source: www.statista.com/statistics/267161/market-share-of-search-engines-in-the-united-states/.

very fragmented, such as automobile repairs, where small independent shops often dominate and large chains may or may not exist. New products can lead to market concentration: It is estimated that the Apple iPod had a world market share of over 70 percent among MP3 players in 2009.

THE IMPORTANCE OF MARKET STRUCTURE

Consider the evolution of television broadcasting. As the market environment for television broadcasting evolved, the market structure changed, resulting in a new set of challenges and choices. In the early days, there was only one choice: the "free" analog channels that were broadcast over the airwaves. In most countries, there was only one channel, owned and run by the government. In the United States, some of the more populated markets were able to receive more channels because local channels were set up to cover a market with more potential viewers. By the 1970s, new technologies made it possible to broadcast by way of cable connectivity and the choices offered to consumers began to expand rapidly. Cable television challenged the "free" broadcast channels by offering more choice and a better-quality picture. The innovation was expensive for consumers and profitable for the cable companies. By the 1990s, a new alternative began to challenge the existing broadcast and cable systems: satellite television. Satellite providers offered a further expanded set of choices, albeit at a higher price than the free broadcast and cable alternatives. In the early 2000s, satellite television providers lowered their pricing to compete directly with the cable providers.

Today, cable program providers, satellite television providers, and terrestrial digital broadcasters that offer premium and pay-per-view channels compete for customers who are increasingly finding content on the internet and on their mobile devices. Companies like Netflix, Apple, and Amazon offered alternative ways for consumers to access content. By 2018, these companies had moved beyond the repackaging of existing shows to developing their own content, mirroring the evolution of cable channels such as HBO and ESPN a decade earlier.

This is a simple illustration of the importance of market structure. As the market for television broadcasting became increasingly competitive, managers have had to make decisions regarding product packaging, pricing, advertising, and marketing in order to survive in the changing environment. In addition, mergers and acquisitions as a response to these competitive pressures have changed the essential structure of the industry.

Market structure can be broken down into four distinct categories: perfect competition, monopolistic competition, oligopoly, and monopoly.

We start with the most competitive environment, **perfect competition**. Unlike some economic concepts, perfect competition is not merely an ideal based on assumptions. Perfect competition is a reality—for example, in several commodities markets, where sellers and buyers have a strictly homogeneous product and no single producer is large enough to influence market prices. Perfect competition's characteristics are well recognized and its long-run outcome unavoidable. Profits under the conditions of perfect competition are driven to the required rate of return paid by the entrepreneur to borrow capital from investors (so-called normal profit or rental cost of capital). This does not mean that all perfectly competitive industries are doomed to extinction by a lack of profits. On the contrary, millions of businesses that do very well are living under the pressures of perfect competition.

Monopolistic competition is also highly competitive; however, it is considered a form of imperfect competition. Two economists, Edward H. Chamberlin (US) and Joan Robinson (UK), identified this hybrid market and came up with the term because there are not only strong elements of competition in this market structure but also some monopoly-like conditions. The competitive characteristic is a notably large number of firms, while the monopoly aspect is the result of product differentiation. That is, if the seller can convince consumers that its product is uniquely different from

other, similar products, then the seller can exercise some degree of pricing power over the market. A good example is the brand loyalty associated with soft drinks such as Coca-Cola. Many of Coca-Cola's customers believe that their beverages are truly different from and better than all other soft drinks. The same is true for fashion creations and cosmetics.

The **oligopoly** market structure is based on a relatively small number of firms supplying the market. The small number of firms in the market means that each firm must consider what retaliatory strategies the other firms will pursue when prices and production levels change. Consider the pricing behavior of commercial airline companies. Pricing strategies and route scheduling are based on the expected reaction of the other carriers in similar markets. For any given route—say, from Paris, France, to Chennai, India—only a few carriers are in competition. If one of the carriers changes its pricing package, others will likely retaliate. Understanding the market structure of oligopoly markets can help in identifying a logical pattern of strategic price changes for the competing firms.

Finally, the least competitive market structure is **monopoly**. In pure monopoly markets, there are no other good substitutes for the given product or service. There is a single seller, which, if allowed to operate without constraint, exercises considerable power over pricing and output decisions. In most market-based economies around the globe, pure monopolies are regulated by a governmental authority. The most common example of a regulated monopoly is the local electrical power provider. In most cases, the monopoly power provider is allowed to earn a normal return on its investment and prices are set by the regulatory authority to allow that return.

2.2 Factors That Determine Market Structure

Five factors determine market structure:

1. The number and relative size of firms supplying the product;
2. The degree of product differentiation;
3. The power of the seller over pricing decisions;
4. The relative strength of the barriers to market entry and exit; and
5. The degree of non-price competition.

The number and relative size of firms in a market influence market structure. If there are many firms, the degree of competition increases. With fewer firms supplying a good or service, consumers are limited in their market choices. One extreme case is the monopoly market structure, with only one firm supplying a unique good or service. Another extreme is perfect competition, with many firms supplying a similar product. Finally, an example of relative size is the automobile industry, in which a small number of large international producers (e.g., Volkswagen and Toyota) are the leaders in the global market, and a number of small companies either have market power because they are niche players (e.g., Ferrari or McLaren) or have little market power because of their narrow range of models or limited geographical presence (e.g., Mazda or Fiat-Chrysler).

In the case of monopolistic competition, there are many firms providing products to the market, as with perfect competition. However, one firm's product is differentiated in some way that makes it appear better than similar products from other firms. If a firm is successful in differentiating its product, the differentiation will provide pricing leverage. The more dissimilar the product appears, the more the market will resemble the monopoly market structure. A firm can differentiate its product through aggressive advertising campaigns; frequent styling changes; the linking of its product with other, complementary products; or a host of other methods.

When the market dictates the price based on aggregate supply and demand conditions, the individual firm has no control over pricing. The typical hog farmer in Nebraska and the milk producer in Bavaria are **price takers**. That is, they must accept whatever price the market dictates. This is the case under the market structure of perfect competition. In the case of monopolistic competition, the success of product differentiation determines the degree with which the firm can influence price. In the case of oligopoly, there are so few firms in the market that price control becomes possible. However, the small number of firms in an oligopoly market invites complex pricing strategies. Collusion, price leadership by dominant firms, and other pricing strategies can result.

The degree to which one market structure can evolve into another and the difference between potential short-run outcomes and long-run equilibrium conditions depend on the strength of the barriers to entry and the possibility that firms fail to recoup their original costs or lose money for an extended period of time and are therefore forced to exit the market. Barriers to entry can result from very large capital investment requirements, as in the case of petroleum refining. Barriers may also result from patents, as in the case of some electronic products and drug formulas. Another entry consideration is the possibility of high exit costs. For example, plants that are specific to a special line of products, such as aluminum smelting plants, are non-redeployable, and exit costs would be high without a liquid market for the firm's assets. High exit costs deter entry and are therefore also considered barriers to entry. In the case of farming, the barriers to entry are low. Production of corn, soybeans, wheat, tomatoes, and other produce is an easy process to replicate; therefore, those are highly competitive markets.

Non-price competition dominates those market structures where product differentiation is critical. Therefore, monopolistic competition relies on competitive strategies that may not include pricing changes. An example of non-price competition is product differentiation through marketing. In other circumstances, non-price competition may occur because the few firms in the market feel dependent on each other. Each firm fears retaliatory price changes that would reduce total revenue for all of the firms in the market. Because oligopoly industries have so few firms, each firm feels dependent on the pricing strategies of the others. Therefore, non-price competition becomes a dominant strategy.

Exhibit 1 Characteristics of Market Structure

Market Structure	Number of Sellers	Degree of Product Differentiation	Barriers to Entry	Pricing Power of Firm	Non-price Competition
Perfect competition	Many	Homogeneous/ Standardized	Very Low	None	None
Monopolistic competition	Many	Differentiated	Low	Some	Advertising and Product Differentiation
Oligopoly	Few	Homogeneous/ Standardized	High	Some or Considerable	Advertising and Product Differentiation
Monopoly	One	Unique Product	Very High	Considerable	Advertising

From the perspective of the owners of the firm, the most desirable market structure is that with the most control over price, because this control can lead to large profits. Monopoly and oligopoly markets offer the greatest potential control over price;

monopolistic competition offers less control. Firms operating under perfectly competitive market conditions have no control over price. From the consumers' perspective, the most desirable market structure is that with the greatest degree of competition because prices are generally lower. Thus, consumers would prefer as many goods and services as possible to be offered in competitive markets.

As often happens in economics, there is a trade-off. While perfect competition gives the largest quantity of a good at the lowest price, other market forms may spur more innovation. Specifically, there may be high costs in researching a new product, and firms will incur such costs only if they expect to earn an attractive return on their research investment. This is the case often made for medical innovations, for example—the cost of clinical trials and experiments to create new medicines would bankrupt perfectly competitive firms but may be acceptable in an oligopoly market structure. Therefore, consumers can benefit from less-than-perfectly-competitive markets.

PORTER'S FIVE FORCES AND MARKET STRUCTURE

A financial analyst aiming to establish market conditions and consequent profitability of incumbent firms should start with the questions framed by Exhibit 1: How many sellers are there? Is the product differentiated? and so on. Moreover, in the case of monopolies and quasi monopolies, the analyst should evaluate the legislative and regulatory framework: Can the company set prices freely, or are there governmental controls? Finally, the analyst should consider the threat of competition from potential entrants.

This analysis is often summarized by students of corporate strategy as "Porter's five forces," named after Harvard Business School professor Michael E. Porter. His book, *Competitive Strategy*, presented a systematic analysis of the practice of market strategy. Porter (2008) identified the five forces as:

- Threat of entry;
- Power of suppliers;
- Power of buyers (customers);
- Threat of substitutes; and
- Rivalry among existing competitors.

It is easy to note the parallels between four of these five forces and the columns in Exhibit 1. The only "orphan" is the power of suppliers, which is not at the core of the theoretical economic analysis of competition, but which has substantial weight in the practical analysis of competition and profitability.

Some stock analysts (e.g., Dorsey 2004) use the term "economic moat" to suggest that there are factors protecting the profitability of a firm that are similar to the moats (ditches full of water) that were used to protect some medieval castles. A deep moat means that there is little or no threat of entry by invaders, i.e. competitors. It also means that customers are locked in because of high switching costs.

3 PERFECT COMPETITION

Perfect competition is characterized by the five conditions presented in Exhibit 1, above:

1 There are a large number of potential buyers and sellers.
2 The products offered by the sellers are virtually identical.
3 There are few or easily surmountable barriers to entry and exit.

4 Sellers have no market-pricing power.

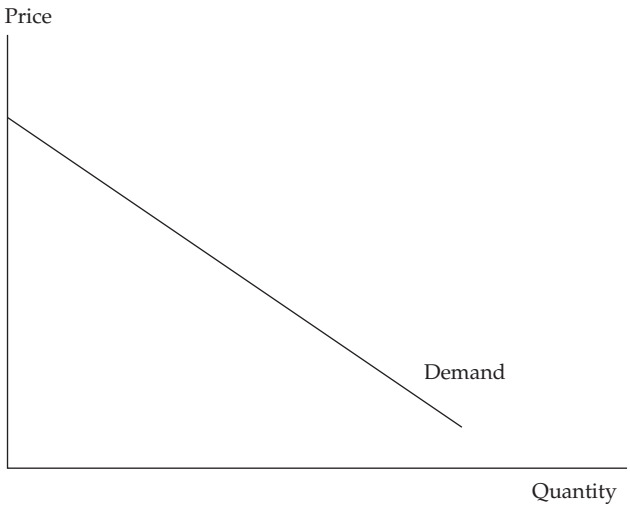
5 Non-price competition is absent.

While few markets achieve the distinction of being perfectly competitive, it is useful to establish the outcome associated with this market structure as a benchmark against which other market structures can be compared. The most typical example of perfect competition is found in certain aspects of the agriculture industry, such as the large number of farmers growing corn for animal feed. Corn is a primary source of food for pork, beef, and poultry production. A bushel of corn from Farmer Brown is virtually identical to a bushel of corn from Farmer Lopez. If a hog farmer needs corn to feed his hogs, it does not matter whether the corn comes from Farmer Brown or Farmer Lopez. Furthermore, the aggregate corn market is well defined, with active futures and spot markets. Information about the corn market is easy and inexpensive to access, and there is no way to differentiate the product, such as by advertising. Agribusiness is capital intensive, but where arable land is relatively abundant and water is available, the barriers to entry (e.g., capital and expertise) for corn production are relatively low.

3.1 Demand Analysis in Perfectly Competitive Markets

The price of a homogeneous product sold in a competitive market is determined by the demand and supply in that market. Economists usually represent demand and supply in a market through demand and supply curves in a two-axis plane, where quantity and price are shown on the x-axis and y-axis, respectively. Economists believe that demand functions have negative slopes, as shown in Exhibit 2. That is, at high prices, less is demanded. For normal goods and services, as the price declines, the quantity demanded increases. This concept is based on two effects: the income effect and the substitution effect. The income effect results from the increased purchasing power the consumer has when prices fall. With lower prices, the consumer can afford to purchase more of the product. The substitution effect comes from the increasing attractiveness of the lower-priced product. If soybean prices are unchanged and corn prices decrease, hog farmers will substitute corn for soybeans as feed for their animals.

Exhibit 2 Market Demand in Perfect Competition



Assume the demand for this product can be specified as

$Q_D = 50 - 2P$

where Q_D is the quantity of demand and P is the product's price. This demand function can be rearranged in terms of price:

$$P = 25 - 0.5Q_D$$

In this form, total revenue (TR) is equal to price times quantity, or $P \times Q_D$. Thus,

$$\text{TR} = PQ_D = 25Q_D - 0.5Q_D^2$$

Average revenue (AR) can be found by dividing TR by Q_D. Therefore,

$$\text{AR} = \text{TR}/Q_D = \left(25Q_D - 0.5Q_D^2\right)/Q_D = 25 - 0.5Q_D$$

Note that the AR function is identical to the market demand function. The assumption here is that the relationship between price and quantity demanded is linear. Clearly, that may not be the case in the real market. Another simplifying assumption made is that the price of the product is the only determinant of demand. Again, that is not likely in the real market. For example, economic theory suggests that consumer income is another important factor in determining demand. The prices of related goods and services, such as substitutes and complements, are also considered factors affecting demand for a specific product.

Marginal revenue (MR) is the change in total revenue per extra increment sold when the quantity sold changes by a small increment, ΔQ_D. Substituting $(Q_D + \Delta Q_D)$ into the total revenue (TR) equation, marginal revenue can be expressed as:

$$\text{MR} = \frac{\Delta \text{TR}}{\Delta Q_D} = \frac{\left[25\left(Q_D + \Delta Q_D\right) - 0.5\left(Q_D^2 + 2Q_D\Delta Q_D + \Delta Q_D^2\right)\right] - \left[25Q_D - 0.5Q_D^2\right]}{\Delta Q_D}$$

$$= \frac{25\Delta Q_D - Q_D\Delta Q_D - 0.5\Delta Q_D^2}{\Delta Q_D} = 25 - Q_D - 0.5\Delta Q_D$$

For example, suppose $Q_D = 5$ and $\Delta Q_D = 1$, then total revenue increases from 112.50 [$= 25(5) - 0.5\,(5^2)$] to 132 [$= 25(6) - 0.5(6^2)$], and marginal revenue is 19.5 = (132 – 112.5)/1. Note that marginal revenue is equal to $(25 - Q_D - 0.5\Delta\, Q_D)$. Now suppose that ΔQ_D is much smaller, for example $\Delta Q_D = 0.1$. In this case, total revenue increases to 114.495 [$= 25(5.1) - 0.5(5.1^2)$], and marginal revenue is 1.995/0.1 = 19.95. It is straightforward to confirm that as $\Delta\, Q_D$ gets smaller marginal revenue gets closer to $20 = 25 - Q_D$. So, for very small changes in the quantity sold we can write marginal revenue as[3]

$$\text{MR} = 25 - Q_D$$

Although we have introduced the concept of marginal revenue in the context of the demand curve for the market as a whole, its usefulness derives from its role in the output and pricing decisions of individual firms. As we will see, marginal revenue and an analogous concept, marginal cost, are critical in determining firms' profit-maximizing strategies.

3.1.1 *Elasticity of Demand*

Consumers respond differently to changes in the price of different kinds of products and services. The quantity demanded for some products is very price sensitive, while for other products, price changes result in little change in the quantity demanded. Economists refer to the relationship between changes in price and changes in the quantity demanded as the price elasticity of demand. Therefore, the demand for the former group of products—those that are very price sensitive—is said to have high

3 Readers who are familiar with calculus will recognize this as the derivative of total revenue with respect to the quantity sold.

price elasticity, whereas the demand for the latter group is said to have low price elasticity. Understanding the sensitivity of demand changes to changes in price is critical to understanding market structures.

Price elasticity of demand measures the percentage change in the quantity demanded given a percentage change in the price of a given product. Because the relationship of demand to price is negative, the price elasticity of demand would be negative. *Many economists, however, present the price elasticity as an absolute value, so that price elasticity has a positive sign. We will follow that convention.* Higher price elasticity indicates that consumers are very responsive to changes in price. Lower values for price elasticity imply that consumers are not very responsive to price changes. Price elasticity can be measured with the following relationship:

$$\varepsilon_P = -(\%\text{ change in } Q_D) \div (\%\text{ change in } P)$$

where ε_P is price elasticity of demand, Q_D is the quantity demanded, and P is the product's price.

Price elasticity of demand falls into three categories. When demand is very responsive to price change, it is identified as *elastic*. When demand is not responsive to price change, it is identified as *inelastic*. When the percentage change in quantity demanded is exactly the same as the percentage change in price, the demand is called *unitary elastic*.

$\varepsilon_P > 1$ Demand is elastic

$\varepsilon_P = 1$ Demand is unitary elastic

$\varepsilon_P < 1$ Demand is inelastic

Price elasticity of demand depends on several factors. *Price elasticity will be higher if there are many close substitutes for the product.* If a product has many good alternatives, consumers will be more sensitive to price changes. For example, carbonated beverages ("soft drinks") have many close substitutes. It takes strong brand loyalty to keep customer demand high in the soft drink market when one brand's price is strategically lowered; the price elasticity of demand for Coca-Cola has been estimated to be 3.8. For products with numerous close substitutes, demand is highly elastic. For products with few close substitutes, demand is lower in price elasticity and would be considered price inelastic. The demand for first-class airline tickets is often seen as inelastic because only very wealthy people are expected to buy them; the demand for economy-class tickets is elastic because the typical consumer for this product is more budget-conscious. Consumers do not consider economy-class airline tickets a close substitute for first-class accommodations, particularly on long flights.

The airline ticket example introduces another determinant of price elasticity of demand. *The greater the share of the consumer's budget spent on the item, the higher the price elasticity of demand*. Expensive items, such as durable goods (e.g., refrigerators and televisions), tend to have higher elasticity measures, while less expensive items, such as potatoes and salt, have lower elasticity values. Consumers will not change their normal salt consumption if the price of salt decreases by 10 percent. Instead, they will buy their next package of salt when they run out, with very little regard to the price change.

The airline ticket also makes a good example for the final factor determining price elasticity. *Price elasticity of demand also depends on the length of time within which the demand schedule is being considered*. Holiday airline travel is highly price elastic. Consumers shop vigorously for vacation flights because they have time to plan their holiday. Business airline travelers typically have less flexibility in determining their schedules. If your business requires a face-to-face meeting with a client, then the price of the ticket is somewhat irrelevant. If gasoline prices increase, there is very little you can do in the short run but pay the higher price. However, evidence of commuter choices indicates that many use alternative transportation methods after the gasoline

price spikes. In the long run, higher gasoline prices will lead consumers to change their modes of transportation, trading in less efficient vehicles for automobiles with higher gas mileage or public transit options where available.

There are two extreme cases of price elasticity of demand. One extreme is the **horizontal demand schedule**. This term implies that at a given price, the response in the quantity demanded is infinite. *This is the demand schedule faced by a perfectly competitive firm because it is a price taker*, as in the case of a corn farmer. If the corn farmer tried to charge a higher price than the market price, nobody would buy her product. On the other hand, the farmer has no incentive to sell at a lower price because she can sell all she can produce at the market price. In a perfectly competitive market the quantity supplied by an individual firm has a negligible effect on the market price. In the case of *perfect price elasticity*, the measure is $\varepsilon_P = \infty$.

The other extreme is the **vertical demand schedule**. The vertical demand schedule implies that some fixed quantity is demanded, regardless of price. An example of such demand is the diabetic consumer with the need for a certain amount of insulin. If the price of insulin goes up, the patient will not consume less of it. The amount desired is set by the patient's medical condition. The measure for *perfect price inelasticity* is $\varepsilon_P = 0$.

The nature of the elasticity calculation and consumer behavior in the marketplace imply that for virtually any product (excluding cases of perfect elasticity and perfect inelasticity), demand is more elastic at higher prices and less elastic (more inelastic) at lower prices. For example, at current low prices, the demand for table salt is very inelastic. However, if table salt increased in price to hundreds of dollars per ounce, consumers would become more responsive to its price changes. Exhibit 3 reports several empirical estimates of price elasticity of demand.

Exhibit 3 Empirical Price Elasticities[4]

Commodity (Good/Service)	Price Elasticity of Market Demand
Alcoholic beverages consumed at home	
Beer	0.84
Wine	0.55
Liquor	0.50
Coffee	
Regular	0.16
Instant	0.36
Credit charges on bank cards	2.44
Furniture	3.04
Glassware/china	1.20
International air transportation United States/Europe	1.20
Shoes	0.73
Soybean meal	1.65
Tomatoes	2.22

4 Various sources, as noted in McGuigan, Moyer, and Harris (2008), p. 95. These are the elasticities with respect to the product's own price; by convention, they are shown here as positive numbers.

3.1.2 *Other Factors Affecting Demand*

There are two other important forces that influence shifts in consumer demand. One influential factor is consumer income and the other is the price of a related product. For normal goods, as consumer income increases, the demand increases. The degree to which consumers respond to higher incomes by increasing their demand for goods and services is referred to as income elasticity of demand. **Income elasticity of demand** measures the responsiveness of demand to changes in income. The calculation is similar to that of price elasticity, with the percentage change in income replacing the percentage change in price. Note the new calculation below:

$$\varepsilon_Y = (\%\text{ change in } Q_D) \div (\%\text{ change in } Y)$$

where ε_Y is income elasticity of demand, Q_D is the quantity demanded, and Y is consumer income. For normal goods, the measure ε_Y will be a positive value. That is, as consumers' income rises, more of the product is demanded. For products that are considered luxury items, the measure of income elasticity will be greater than one. There are other goods and services that are considered inferior products. For inferior products, as consumer income rises, less of the product is demanded. Inferior products will have negative values for income elasticity. For example, a person on a small income may watch television shows, but if this person had more income, she would prefer going to live concerts and theater performances; in this example, television shows would be the inferior good.

As a technical issue, the difference between price elasticity of demand and income elasticity of demand is that the demand adjustment for price elasticity represents a movement *along the demand schedule* because the demand schedule represents combinations of price and quantity. The demand adjustment for income elasticity represents a *shift in the demand curve* because with a higher income one can afford to purchase more of the good at any price. For a normal good, an increase in income would shift the demand schedule out to the right, away from the origin of the graph, and a decrease in income would shift the demand curve to the left, toward the origin.

The final factor influencing demand for a product is the change in price of a related product, such as a strong substitute or a complementary product. If a close competitor in the beverage market lowers its price, consumers will substitute that product for your product. Thus, your product's demand curve will shift to the left, toward the origin of the graph. **Cross-price elasticity of demand** is the responsiveness of the demand for product A that is associated with the change in price of product B:

$$\varepsilon_X = (\%\text{ change in } Q_{DA}) \div (\%\text{ change in } P_B)$$

where ε_X is cross-price elasticity of demand, Q_{DA} is the quantity demanded of product A, and P_B is the price of product B.

When the cross-price elasticity of demand between two products is *positive*, the two products are considered to be **substitutes**. For example, you may expect to have positive cross-price elasticity between honey and sugar. If the measure of cross-price elasticity is *negative*, the two products are referred to as **complements** of each other. For example, if the price of DVDs goes up, you would expect consumers to buy fewer DVD players. In this case, the cross-price elasticity of demand would have a negative value.

Reviewing cross-price elasticity values provides a simple test for the degree of competition in the market. The more numerous and the closer the substitutes for a product, the lower the pricing power of firms selling in that market; the fewer the substitutes for a product, the greater the pricing power. One interesting application was a US Supreme Court case involving the production and sale of cellophane by DuPont.[5]

5 *US v. DuPont*, 351 US 377 (1956), as noted in McGuigan, Moyer, and Harris (2008).

The court noted that the relevant product market for DuPont's cellophane was the broader flexible packaging materials market. The Supreme Court found the cross-price elasticity of demand between cellophane and other flexible packaging materials to be sufficiently high and exonerated DuPont from a charge of monopolizing the market.

Because price elasticity of demand relates changes in price to changes in the quantity demanded, there must be a logical relationship between marginal revenue and price elasticity. Recall that marginal revenue equals the change in total revenue given a change in output or sales. An increase in total revenue results from a decrease in price that results in an increase in sales. In order for the increase in the quantity demanded to be sufficient to offset the decline in price, the percentage change in quantity demanded must be greater than the percentage decrease in price. The relationship between TR and price elasticity is as follows:

$\varepsilon_P > 1$ Demand is elastic	$\uparrow P \rightarrow$ TR $\downarrow$ and $\downarrow P \rightarrow$ TR $\uparrow$
$\varepsilon_P = 1$ Demand is unitary elastic	$\updownarrow P \rightarrow$ no change in TR
$0 < \varepsilon_P < 1$ Demand is inelastic	$\uparrow P \rightarrow$ TR $\uparrow$ and $\downarrow P \rightarrow$ TR $\downarrow$

Total revenue is maximized when marginal revenue is zero. The logic is that as long as marginal revenue is positive (i.e., each additional unit sold contributes to additional total revenue), total revenue will continue to increase. Only when marginal revenue becomes negative will total revenue begin to decline. Therefore, the percentage decrease in price is greater than the percentage increase in quantity demanded. The relationship between marginal revenue (MR) and price elasticity can be expressed as

$$\text{MR} = P[1 - (1/\varepsilon_P)]$$

An understanding of price elasticity of demand is an important strategic tool. It would be very useful to know in advance what would happen to your firm's total revenue if you increased the product's price. If you are operating in the inelastic portion of the demand curve, increasing the price of the product will increase total revenue. On the other hand, if you are operating in the elastic portion of the product's demand curve, increasing the price will decrease total revenue.

Decision makers can also use the relationship between marginal revenue and price elasticity of demand in other ways. For example, suppose you are a farmer considering planting soybeans or some other feed crop, such as corn. From Exhibit 3, we know that soybean meal's price elasticity of demand has been estimated to be 1.65. We also know that the current (May 2018) soybean meal price is $465.00 per metric ton.[6] Therefore, by solving the equation above, we find that the expected marginal revenue per metric ton of soybean meal is $183.16. Soybeans may prove to be a profitable crop for the farmer. Just a few years earlier, in May of 2014, the price of a metric ton of soybean meal was $578.75. Given the crop's price elasticity of demand, the estimated marginal revenue per metric ton was then $227.97. The higher price translates into higher marginal revenue and might have induced the farmer to plant even more soybeans rather than another feed crop instead.

How do business decision makers decide what level of output to bring to the market? To answer that question, the firm must understand its cost of resources, its production relations, and its supply function. Once the supply function is well defined and understood, it is combined with the demand analysis to determine the profit-maximizing levels of output.

6 Source: World-Bank Commodity Market Report 2018.

3.1.3 *Consumer Surplus: Value Minus Expenditure*

To this point, we have discussed the fundamentals of supply and demand curves and explained a simple model of how a market can be expected to arrive at an equilibrium combination of price and quantity. While it is certainly necessary for the analyst to understand the basic workings of the market model, it is also crucial to have a sense of why we might care about the nature of the equilibrium. In this section we review the concept of **consumer surplus,** which is helpful in understanding and evaluating business pricing strategies. Consumer surplus is defined as the difference between the value that a consumer places on the units purchased and the amount of money that was required to pay for them. It is a measure of the value gained by the buyer from the transaction.

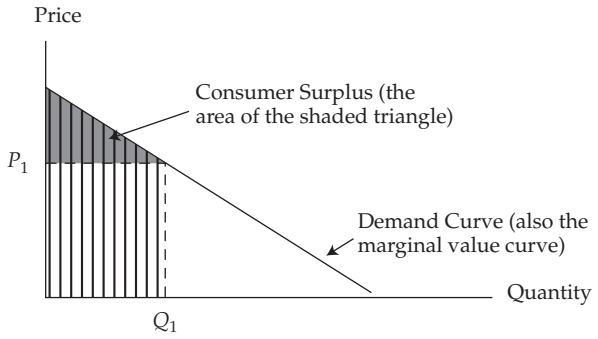
To get an intuitive feel for the concept of consumer surplus, consider the last thing you purchased. Whatever it was, think of how much you paid for it. Now contrast that price with the maximum amount you *would have been willing to pay* rather than go without the item altogether. If those two numbers are different, we say you received some consumer surplus from your purchase. You got a "bargain" because you would have been willing to pay more than you had to pay.

Earlier, we referred to the law of demand, which says that as price falls, consumers are willing to buy more of the good. This observation translates into a negatively sloped demand curve. Alternatively, we could say that the highest price that consumers are willing to pay for an additional unit declines as they consume more and more of a good. In this way, we can interpret their *willingness to pay* as a measure of how much they *value* each additional unit of the good. This is a very important point: To purchase a unit of some good, consumers must give up something else they value. So, the price they are willing to pay for an additional unit of a good is a measure of how much they value that unit, in terms of the other goods they must sacrifice to consume it.

If demand curves are negatively sloped, it must be because the value of each additional unit of the good falls as more of the good is consumed. We shall explore this concept further below, but for now, it is enough to recognize that the demand curve can therefore be considered a **marginal value curve**, because it shows the highest price consumers would be willing to pay for each additional unit. In effect, the demand curve is the willingness of consumers to pay for each additional unit.

This interpretation of the demand curve allows us to measure the total value of consuming any given quantity of a good: It is the sum of all the marginal values of each unit consumed, up to and including the last unit. Graphically, this measure translates into the area under the consumer's demand curve, up to and including the last unit consumed, as shown in Exhibit 4, where the consumer is choosing to buy Q_1 units of the good at a price of P_1. The marginal value of the $Q_1{}^{\text{th}}$ unit is clearly P_1 because that is the highest price the consumer is willing to pay for that unit. Importantly, however, the marginal value of each unit *up to* the $Q_1{}^{\text{th}}$ is greater than P_1.

Exhibit 4 Consumer Surplus



Note: Consumer surplus is the area beneath the demand curve and above the price paid.

Because the consumer would have been willing to pay more for each of those units than she paid (P_1), we can say she received more value than the cost to her of buying them. This extra value is the buyer's consumer surplus. The *total value* of quantity Q_1 to the buyer is the area of the vertically crosshatched trapezoid in Exhibit 4. The *total expenditure* is only the area of the rectangle with height P_1 and base Q_1 (bottom section). The total consumer surplus received from buying Q_1 units at a level price of P_1 per unit is the difference between the area under the demand curve and the area of the rectangle $P_1 \times Q_1$. The resulting area is shown as the lightly shaded triangle (upper section).

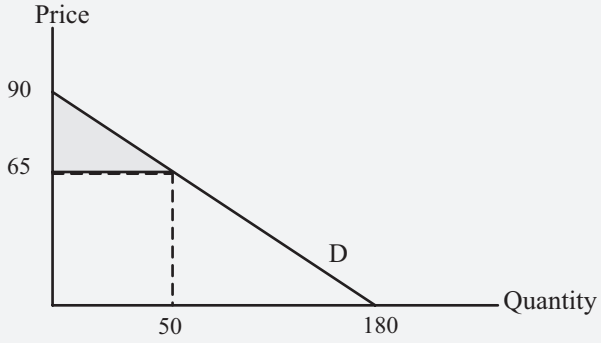
EXAMPLE 1

Consumer Surplus

A market demand function is given by the equation $Q_D = 180 - 2P$. Find the value of consumer surplus if price is equal to 65.

Solution:

First, input 65 into the demand function to find the quantity demanded at that price: $Q_D = 180 - 2(65) = 50$. Then, to make drawing the demand curve easier, invert the demand function by solving for P in terms of Q_D: $P = 90 - 0.5Q_D$. Note that the price intercept is 90 and the quantity intercept is 180. Draw the demand curve:

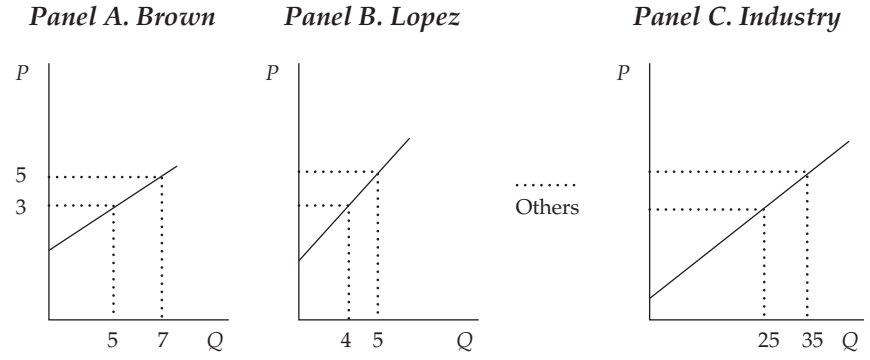


Find the area of the triangle above the price of 65 and below the demand curve, up to quantity 50: Area = ½ (Base)(Height) = ½ (50)(25) = 625.

3.2 Supply Analysis in Perfectly Competitive Markets

Consider two corn farmers, Mr. Brown and Ms. Lopez. They both have land available to them to grow corn and can sell at one price, say 3 currency units per kilogram. They will try to produce as much corn as is profitable at that price. If the price is driven up to 5 currency units per kilogram by new consumers entering the market—say, ethanol producers—Mr. Brown and Ms. Lopez will try to produce more corn. To increase their output levels, they may have to use less productive land, increase irrigation, use more fertilizer, or all three. Their production costs will likely increase. They will both still try to produce as much corn as possible to profit at the new, higher price of 5 currency units per kilogram. Exhibit 5 illustrates this example. Note that the supply functions for the individual firms have positive slopes. Thus, as prices increase, the firms supply greater quantities of the product.

Exhibit 5 Firm and Market Supply in Perfect Competition



Notice that the market supply curve is the sum of the supply curves of the individual firms—Brown, Lopez, and others—that make up the market. Assume that the supply function for the market can be expressed as a linear relationship, as follows:

$Q_S = 10 + 5P$, or $P = -2 + 0.2Q_S$,

where Q_S is the quantity supplied and P is the price of the product.

Before we analyze the optimal supply level for the firm, we need to point out that economic costs and profits differ from accounting costs and profits in a significant way. **Economic costs** include all the remuneration needed to keep the productive resource in its current employment or to acquire the resource for productive use.

To evaluate the remuneration needed to keep the resource in its current use and attract new resources for productive use, economists refer to the resource's **opportunity cost**. Opportunity cost is measured by determining the resource's next best opportunity. If a corn farmer could be employed in an alternative position in the labor market with an income of 50,000, then the opportunity cost of the farmer's labor is 50,000. Similarly, the farmer's land and capital could be leased to another farmer or sold and reinvested in another type of business. The return foregone by not doing so is an opportunity cost. In economic terms, total cost includes the full normal market return on all the resources utilized in the business. **Economic profit** is the difference between TR and total cost (TC). Economic profit differs from accounting profit because accounting profit does not include opportunity cost. Accounting profit includes only explicit payments to outside providers of resources (e.g. workers, vendors, lenders) and depreciation based on the historic cost of physical capital.

3.3 Optimal Price and Output in Perfectly Competitive Markets

Carrying forward our examples from Sections 3.1 and 3.2, we can now combine the market supply and demand functions to solve for the equilibrium price and quantity, where Q^* represents the equilibrium level of both supply and demand.

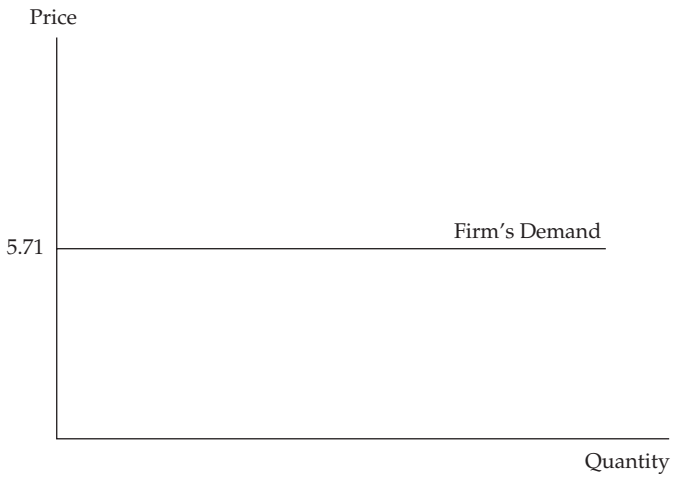
$$\begin{aligned} P = 25 - 0.5Q_D &= -2 + 0.2Q_S = P \\ 25 - 0.5Q_D &= -2 + 0.2Q_S \\ 27 &= 0.7Q^* \\ Q^* &= 38.57 \end{aligned}$$

According to the market demand curve, the equilibrium price is

$$P = 25 - 0.5Q^* = 25 - 0.5(38.57) = 25 - 19.29 = 5.71.$$

With many firms in the market and total output in the market of almost 39 units of the product, the effective market price would be 5.71. This result becomes the demand function for each perfectly competitive firm. Even if a few individual producers could expand production, there would not be a noticeable change in the market equilibrium price. In fact, if any one firm could change the equilibrium market price, the market would not be in perfect competition. Therefore, the demand curve that each perfectly competitive firm faces is a horizontal line at the equilibrium price, as shown in Exhibit 6, even though the demand curve for the whole market is downward sloping.

Exhibit 6 Individual Firm's Demand in Perfect Competition



EXAMPLE 2

Demand Curves in Perfect Competition

Is it possible that the demand schedule faced by Firm A is horizontal while the demand schedule faced by the market as a whole is downward sloping?

A No, because Firm A can change its output based on demand changes.

B No, because a horizontal demand curve means that elasticity is infinite.

C Yes, because consumers can go to another firm if Firm A charges a higher price, and Firm A can sell all it produces at the market price.

Solution:

C is correct. Firm A cannot charge a higher price and has no incentive to sell at a price below the market price.

To analyze the firm's revenue position, recall that average revenue is equivalent to the firm's demand function. Therefore, the horizontal line that represents the firm's demand curve is the firm's AR schedule.

Marginal revenue is the incremental increase in total revenue associated with each additional unit sold. For every extra unit the firm sells, it receives 5.71. Thus, the firm's MR schedule is also the horizontal line at 5.71. TR is calculated by multiplying AR by the quantity of products sold. Total revenue is the area under the AR line at the point where the firm produces the output. In the case of perfect competition, the following conditions hold for the individual firm:

Price = Average revenue = Marginal revenue

The next step is to develop the firm's cost functions. The firm knows that it can sell the entire product it produces at the market's equilibrium price. How much should it produce? That decision is determined by analysis of the firm's costs and revenues. A corn farmer uses three primary resources: land, labor, and capital. In economics, capital is any man-made aid to production. For the corn farmer, his or her capital includes the irrigation system, tractors, harvesters, trucks, grain bins, fertilizer, and so forth. The labor includes the farmer, perhaps members of the farmer's family, and hired labor. In the initial stages of production, only the farmer and the farmer's family are cultivating the land, with a significant investment in capital. They have a tractor, fertilizer, irrigation equipment, grain bins, seed, and a harvester. The investment in land and capital is relatively high compared with the labor input. In this production phase, the average cost of producing a bushel of corn is high. As they begin to expand by adding labor to the collection of expensive land and capital, the average cost of producing corn begins to decline—for example, because one tractor can be used more intensively to plow a larger amount of land. When the combination of land, labor, and capital approaches an efficient range, the average cost of producing a bushel of corn declines.

Given a certain level of technology, there is a limit to the increase in productivity. Eventually something begins to cause declining marginal productivity. That is, each additional unit of input produces a progressively smaller increase in output. This force is called the **law of diminishing returns**. This "law" helps define the shape of the firm's cost functions. Average cost and marginal cost will be U-shaped. Over the initial stages of output, average and marginal costs will decline. At some level of output, the law of diminishing returns will overtake the efficiencies in production and average and marginal costs will increase.

Average cost (AC) is Total cost (TC) divided by Output (Q). Therefore,

$$AC = TC/Q$$

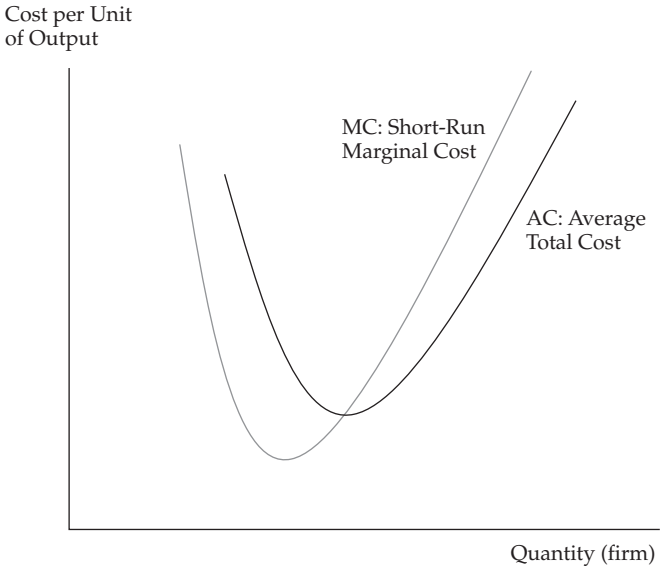
Note that we have defined average cost (AC) in terms of total costs. Many authors refer to this as "average total cost" to distinguish it from a related concept, "average variable cost," which omits fixed costs. In the remainder of this reading, *average cost should be understood to mean average total cost.*

Marginal cost (MC) is the change in TC associated with an incremental change in output:

$$MC = \Delta TC/\Delta Q$$

By definition, fixed costs do not vary with output, so marginal cost reflects only changes in variable costs.[7] MC declines initially because processes can be made more efficient and specialization makes workers more proficient at their tasks. However, at some higher level of output, MC begins to increase (e.g., must pay workers a higher wage to have them work overtime and, in agriculture, less fertile land must be brought into production). MC and AC will be equal at the level of output where AC is minimized. This is a mathematical necessity and intuitive. If you employ the least expensive labor in the initial phase of production, average and marginal cost will decline. Eventually, additional labor will be more costly. For example, if the labor market is at or near full employment, in order to attract additional workers, you must pay higher wages than they are currently earning elsewhere. Thus, the additional (marginal) labor is more costly, and the higher cost increases the overall average as soon as MC exceeds AC. Exhibit 7 illustrates the relationship between AC and MC.

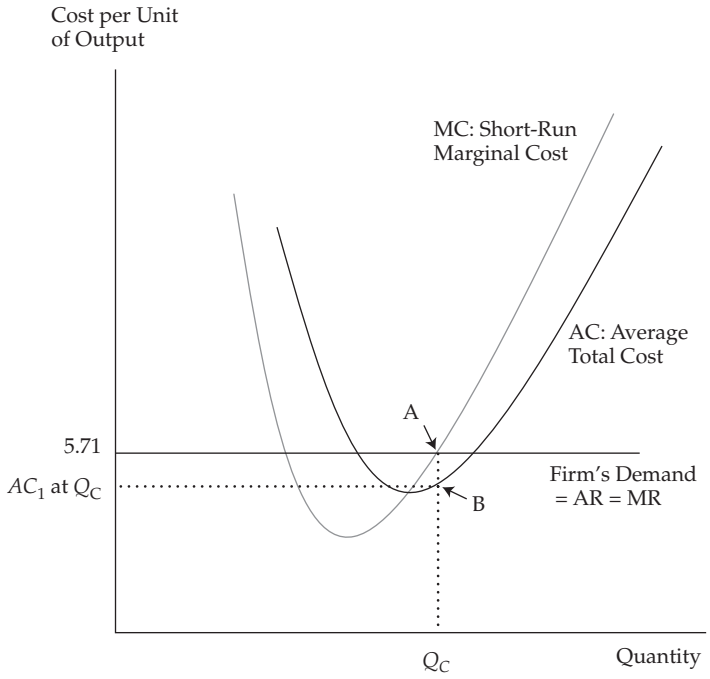
Exhibit 7 Individual Firm's Short-Run Cost Schedules



Now combine the revenue and cost functions from Exhibits 6 and 7. In short-run equilibrium, the perfectly competitive firm can earn an economic profit (or an economic loss). In this example, the equilibrium price, 5.71, is higher than the minimum AC. The firm will always maximize profit at an output level where MR = MC. Recall that in perfect competition, the horizontal demand curve is the marginal revenue and average revenue schedules. By setting output at point A in Exhibit 8, where MR = MC, the firm will maximize profits. Total revenue is equal to $P \times Q$—in this case, 5.71 times Q_C. Total cost is equal to Q_C times the average cost of producing Q_C, at point B in Exhibit 8. The difference between the two areas is economic profit.

7 Readers who are familiar with calculus will recognize that MC is simply the derivative of total cost with respect to quantity produced.

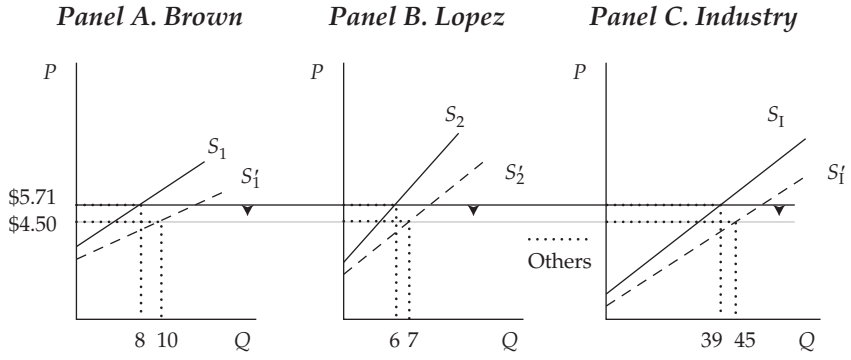
Exhibit 8 Perfectly Competitive Firm's Short-Run Equilibrium



3.4 Factors Affecting Long-Run Equilibrium in Perfectly Competitive Markets

In the long run, economic profit will attract other entrepreneurs to the market, resulting in the production of more output. The aggregate supply will increase, shifting the industry supply (S_1) curve to the right, away from the origin of the graph. For a given demand curve, this increase in supply at each price level will lower the equilibrium price, as shown in Exhibit 9.

Exhibit 9 Perfectly Competitive Market with Increased Supply

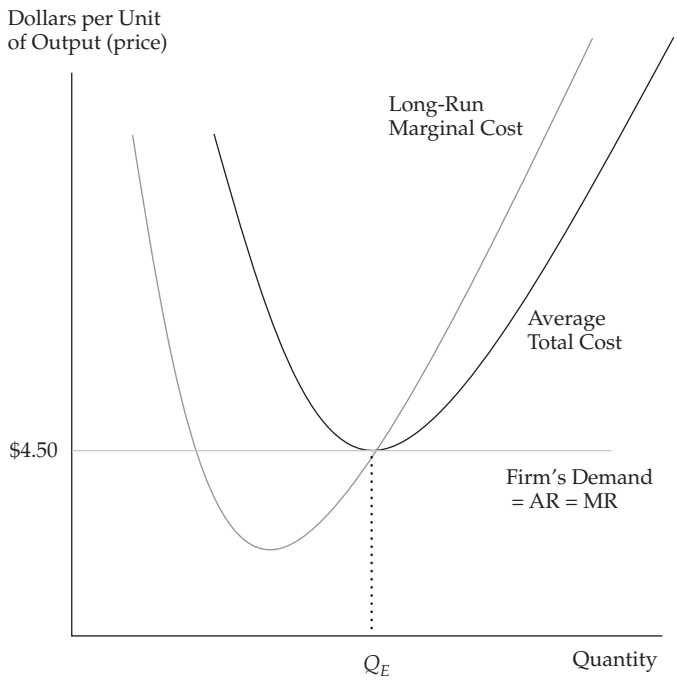


In the long run, the perfectly competitive firm will operate at the point where marginal cost equals the minimum of average cost, because at that point, entry is no longer profitable: In equilibrium, price equals not only marginal cost (firm equilibrium) but also minimum average cost, so that total revenues equal total costs. This

result implies that the perfectly competitive firm operates with zero economic profit. That is, the firm receives its normal profit (rental cost of capital), which is included in its economic costs. Recall that economic profits occur when total revenue exceeds total cost (and therefore differ from accounting profits). With low entry cost and homogeneous products to sell, the perfectly competitive firm earns zero economic profit in the long run.

Exhibit 10 illustrates the long-run equilibrium position of the perfectly competitive firm. Note that total revenue equals price ($4.50) times quantity ($Q_E$) and total cost equals average cost ($4.50) times quantity ($Q_E$).

Exhibit 10 Perfectly Competitive Firm's Long-Run Equilibrium



The long-run marginal cost schedule is the perfectly competitive firm's supply curve. The firm's demand curve is dictated by the aggregate market's equilibrium price. The basic rule of profit maximization is that MR = MC, as is the case in long-run equilibrium. The firm's demand schedule is the same as the firm's marginal revenue and average revenue. Given its cost of operation, the only decision the perfectly competitive firm faces is how much to produce. The answer is the level of output that maximizes its return, and that level is where MR = MC. The demand curve is perfectly elastic. Of course, the firm constantly tries to find ways to lower its cost in the long run.

SCHUMPETER ON INNOVATION AND PERFECT COMPETITION

The Austrian-American economist Joseph A. Schumpeter[8] pointed out that technical change in economics can happen in two main ways:

1 Innovation of process: a new, more efficient way to produce an existing good or service.

2 Innovation of product: a new product altogether or an innovation upon an existing product.

Innovation of process is related to production methods. For example, instead of mixing cement by hand, since the invention of the electric engine it has been possible to use electric mixers. A more recent innovation has been to use the internet to provide technical support to personal computer users: A technician can remotely log on to the customer's PC and fix problems instead of providing instructions over the phone. The result is likely the same, but the process is more efficient.

Innovation of product is related to the product itself. MP3 players, smart phones, robot surgery, and GPS vehicle monitoring have existed only for a few years. They are new products and services. While portable music players existed before the MP3 player, no similar service existed before GPS monitoring of personal vehicles and freight trucks was invented.

How does the reality of continuous innovation of product and process, which is a characteristic of modern economies, fit into the ideal model of perfect competition, where the product is made by a huge number of tiny, anonymous suppliers? This seems a contradiction because the tiny suppliers cannot all be able to invent new products—and indeed, the markets for portable music players and smart phones do not look like perfect competition.

Schumpeter suggested that perfect competition is more of a long-run type of market. In the short run, a company develops a new process or product and is the only one to take advantage of the innovation. This company likely will have high profits and will outpace any competitors. A second stage is what Schumpeter called the swarming (as when a group of bees leaves a hive to follow a queen): In this case, some entrepreneurs notice the innovation and follow the innovator through imitation. Some of them will fail, while others will succeed and possibly be more successful than the initial innovator. The third stage occurs when the new technology is no longer new because everyone has imitated it. At this point, no economic profits are realized, because the new process or product is no longer a competitive advantage, in the sense that everyone has it—which is when perfect competition prevails and we have long-run equilibrium until a new innovation of process or product is introduced.

MONOPOLISTIC COMPETITION

4

Early in the 20th century, economists began to realize that most markets did not operate under the conditions of perfect competition.[9] Many market structures exhibited characteristics of strong competitive forces; however, other distinct non-competitive factors played important roles in the market. As the name implies, monopolistic competition is a hybrid market. *The most distinctive factor in monopolistic competition is product differentiation.* Recall the characteristics from Exhibit 1:

1 There are a large number of potential buyers and sellers.

8 See part 2 of Schumpeter (1942) for the famous "creative destruction" process.
9 Chamberlin (1933).

2 The products offered by each seller are close substitutes for the products offered by other firms, and each firm tries to make its product look different.

3 Entry into and exit from the market are possible with fairly low costs.

4 Firms have some pricing power.

5 Suppliers differentiate their products through advertising and other non-price strategies.

While the market is made up of many firms that compose the product group, each producer attempts to distinguish its product from that of the others. Product differentiation is accomplished in a variety of ways. For example, consider the wide variety of communication devices available today. Decades ago, when each communication market was controlled by a regulated single seller (the telephone company), all telephones were alike. In the deregulated market of today, the variety of physical styles and colors is extensive. All versions accomplish many of the same tasks.

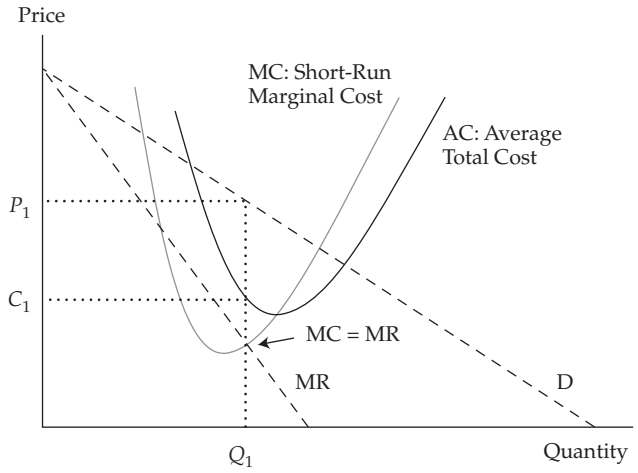
The communication device manufacturers and providers differentiate their products with different colors, styles, networks, bundled applications, conditional contracts, functionality, and more. Advertising is usually the avenue pursued to convince consumers there is a difference between the goods in the product group. Successful advertising and trademark branding result in customer loyalty. A good example is the brand loyalty associated with Harley-Davidson motorcycles. Harley-Davidson's customers believe that their motorcycles are truly different from and better than all other motorcycles. The same kind of brand loyalty exists for many fashion creations and cosmetics.

The extent to which the producer is successful in product differentiation determines pricing power in the market. Very successful differentiation results in a market structure that resembles the single-seller market (monopoly). However, because there are relatively low entry and exit costs, competition will, in the long run, drive prices and revenues down toward an equilibrium similar to perfect competition. Thus, the hybrid market displays characteristics found in both perfectly competitive and monopoly markets.

4.1 Demand Analysis in Monopolistically Competitive Markets

Because each good sold in the product group is somewhat different from the others, the demand curve for each firm in the monopolistic competition market structure is downward sloping to the right. Price and the quantity demanded are negatively related. Lowering the price will increase the quantity demanded, and raising the price will decrease the quantity demanded. There will be ranges of prices within which demand is elastic and (lower) prices at which demand is inelastic. Exhibit 11 illustrates the demand, marginal revenue, and cost structures facing a monopolistically competitive firm in the short run.

Exhibit 11 Short-Run Equilibrium in Monopolistic Competition



In the short run, the profit-maximizing choice is the level of output where MR = MC. Because the product is somewhat different from that of the competitors, the firm can charge the price determined by the demand curve. Therefore, in Exhibit 11, Q_1 is the ideal level of output and P_1 is the price consumers are willing to pay to acquire that quantity. Total revenue is the area of the rectangle $P_1 \times Q_1$.

4.2 Supply Analysis in Monopolistically Competitive Markets

In perfect competition, the firm's supply schedule is represented by the marginal cost schedule. In monopolistic competition, there is no well-defined supply function. The information used to determine the appropriate level of output is based on the intersection of MC and MR. However, the price that will be charged is based on the market demand schedule. The firm's supply curve should measure the quantity the firm is willing to supply at various prices. That information is not represented by either marginal cost or average cost.

4.3 Optimal Price and Output in Monopolistically Competitive Markets

As seen in Section 4.1, in the short run, the profit-maximizing choice is the level of output where MR = MC and total revenue is the area of the rectangle $P_1 \times Q_1$ in Exhibit 11.

The average cost of producing Q_1 units of the product is C_1, and the total cost is the area of the rectangle $C_1 \times Q_1$. The difference between TR and TC is economic profit. The profit relationship is described as

$$\pi = TR - TC$$

where π is total profit, TR is total revenue, and TC is total cost.

THE BENEFITS OF IMPERFECT COMPETITION

Is monopolistic competition indeed imperfect—that is, is it a bad thing? At first, one would say that it is an inefficient market structure because prices are higher and the quantity supplied is less than in perfect competition. At the same time, in the real world, we see

more markets characterized by monopolistic competition than markets meeting the strict conditions of perfect competition. If monopolistic competition were that inefficient, one wonders, why would it be so common?

A part of the explanation goes back to Schumpeter. Firms try to differentiate their products to meet the needs of customers. Differentiation provides a profit incentive to innovate, experiment with new products and services, and potentially improve the standard of living.

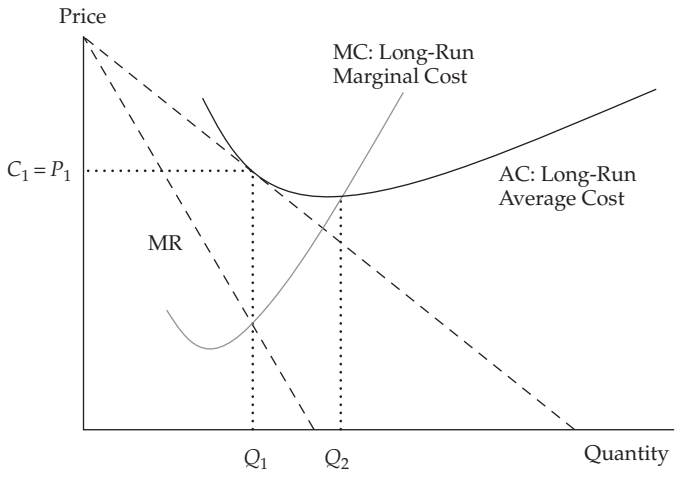
Moreover, because each customer has differing tastes and preferences, slight variations of each good or service are likely to capture the niche of the market that prefers them. An example is the market for candy, where one can find chocolate, licorice, mint, fruit, and many other flavors.

A further reason why monopolistic competition may be good is that people like variety. Traditional economic theories of international trade suggested that countries should buy products from other countries that they cannot produce domestically. Therefore, Norway should buy bananas from a tropical country and sell crude oil in exchange. But this is not the only kind of exchange that happens in reality: For example, Germany imports Honda, Subaru, and Toyota cars from Japan and sells Volkswagen, Porsche, Mercedes, and BMW cars to Japan. In theory, this should not occur because each of the countries produces good cars domestically and does not need to import them. The truth, however (see, for example, Krugman 1989), is that consumers in both countries enjoy variety. Some Japanese drivers prefer to be at the steering wheel of a BMW; others like Hondas, and the same happens in Germany. Variety and product differentiation, therefore, are not necessarily bad things.

4.4 Factors Affecting Long-Run Equilibrium in Monopolistically Competitive Markets

Because TC includes all costs associated with production, including opportunity cost, economic profit is a signal to the market, and that signal will attract more competition. Just as with the perfectly competitive market structure, with relatively low entry costs, more firms will enter the market and lure some customers away from the firm making an economic profit. The loss of customers to new entrant firms will drive down the demand for all firms producing similar products. In the long run for the monopolistically competitive firm, economic profit will fall to zero. Exhibit 12 illustrates the condition of long-run equilibrium for monopolistic competition.

Exhibit 12 Long-Run Equilibrium in Monopolistic Competition



In long-run equilibrium, output is still optimal at the level where MR = MC, which is Q_1 in Exhibit 12. Again, the price consumers are willing to pay for any amount of the product is determined from the demand curve. That price is P_1 for the quantity Q_1 in Exhibit 12, and total revenue is the area of the rectangle $P_1 \times Q_1$. Notice that unlike long-run equilibrium in perfect competition, in the market of monopolistic competition, the equilibrium position is at a higher level of average cost than the level of output that minimizes average cost. Average cost does not reach its minimum until output level Q_2 is achieved. Total cost in this long-run equilibrium position is the area of the rectangle $C_1 \times Q_1$. Economic profit is total revenue minus total cost. In Exhibit 12, economic profit is zero because total revenue equals total cost: $P_1 \times Q_1 = C_1 \times Q_1$.

In the hybrid market of monopolistic competition, zero economic profit in long-run equilibrium resembles perfect competition. However, the long-run level of output, Q_1, is less than Q_2, which corresponds to the minimum average cost of production and would be the long run level of output in a perfectly competitive market. In addition, the economic cost in monopolistic competition includes some cost associated with product differentiation, such as advertising. In perfect competition, there are no costs associated with advertising or marketing because all products are homogeneous. Prices are lower, but consumers may have little variety.

OLIGOPOLY

5

An oligopoly market structure is characterized by only a few firms doing business in a relevant market. The products must all be similar and generally be substitutes for one another. In some oligopoly markets, the goods or services may be differentiated by marketing and strong brand recognition, as in the markets for breakfast cereals and for bottled or canned beverages. Other examples of oligopoly markets are made up of homogeneous products with little or no attempt at product differentiation, such as petroleum and cement. *The most distinctive characteristic of oligopoly markets is the small number of firms that dominate the market. There are so few firms in the relevant market that their pricing decisions are interdependent.* That is, each firm's pricing decision is based on the expected retaliation by the other firms. Recall from Exhibit 1 the characteristics of oligopoly markets:

1. There are a small number of potential sellers.
2. The products offered by each seller are close substitutes for the products offered by other firms and may be differentiated by brand or homogeneous and unbranded.
3. Entry into the market is difficult, with fairly high costs and significant barriers to competition.
4. Firms typically have substantial pricing power.
5. Products are often highly differentiated through marketing, features, and other non-price strategies.

Because there are so few firms, each firm can have some degree of pricing power, which can result in substantial profits. Another by-product of the oligopoly market structure is the attractiveness of price collusion. Even without price collusion, a dominant firm may easily become the price maker in the market. Oligopoly markets without collusion typically have the most sophisticated pricing strategies. Examples of non-colluding oligopolies include the US tobacco market and the Thai beer market.

In 2004, four firms controlled 99 percent of the US tobacco industry.[10] Brands owned by Singha Co. and by ThaiBev controlled over 90 percent of the Thai beer market in 2009. (This situation is expected to change soon, as the Association of Southeast Asian Nations trade agreement will open the doors to competition from other ASEAN producers.) Perhaps the most well-known oligopoly market with collusion is the OPEC cartel, which seeks to control prices in the petroleum market by fostering agreements among oil-producing countries.

5.1 Demand Analysis and Pricing Strategies in Oligopoly Markets

Oligopoly markets' demand curves depend on the degree of pricing interdependence. In a market where collusion is present, the aggregate market demand curve is divided up by the individual production participants. Under non-colluding market conditions, each firm faces an individual demand curve. Furthermore, non-colluding oligopoly market demand characteristics depend on the pricing strategies adopted by the participating firms. There are three basic pricing strategies: pricing interdependence, the Cournot assumption, and the Nash equilibrium.

The first pricing strategy is to assume pricing interdependence among the firms in the oligopoly. A good example of this situation is any market where there are "price wars," such as the commercial airline industry. For example, flying out of their hubs in Atlanta, both Delta Air Lines and AirTran Airways jointly serve several cities. AirTran is a low-cost carrier and typically offers lower fares to destinations out of Atlanta. Delta tends to match the lower fares for those cities also served by AirTran when the departure and arrival times are similar to its own. However, when Delta offers service to the same cities at different time slots, Delta's ticket prices are higher.

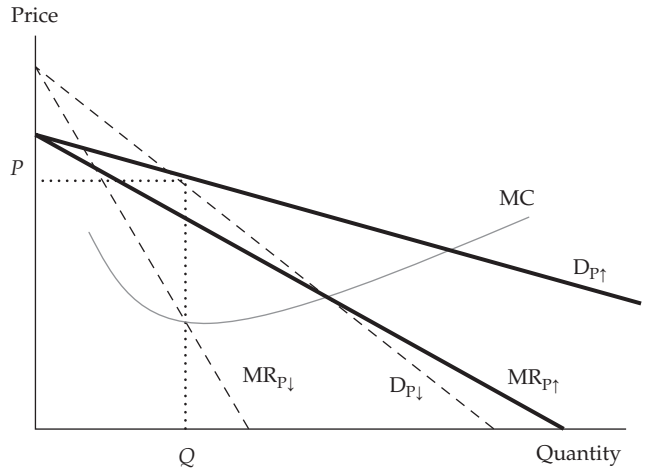
The most common pricing strategy assumption in these price war markets is that competitors will match a price reduction and ignore a price increase. The logic is that by lowering its price to match a competitor's price reduction, the firm will not experience a reduction in customer demand. Conversely, by not matching the price increase, the firm stands to attract customers away from the firm that raised its prices. The oligopolist's demand relationship must represent the potential increase in market share when rivals' price increases are not matched and no significant change in market share when rivals' price decreases are matched.

Given a prevailing price, the price elasticity of demand will be much greater if the price is increased and less if the price is decreased. The firm's customers are more responsive to price increases because its rivals have lower prices. Alternatively, the firm's customers are less responsive to price decreases because its rivals will match its price change.

This implies that the oligopolistic firm faces two different demand structures, one associated with price increases and another relating to price reductions. Each demand function will have its own marginal revenue structure as well. Consider the demand and marginal revenue functions in Exhibit 13(A). The functions $D_{P\uparrow}$ and $MR_{P\uparrow}$ represent the demand and marginal revenue schedules associated with higher prices, while the functions $D_{P\downarrow}$ and $MR_{P\downarrow}$ represent the lower prices' demand and marginal revenue schedules. The two demand schedules intersect at the prevailing price (i.e., the price where price increase and price decrease are both equal to zero).

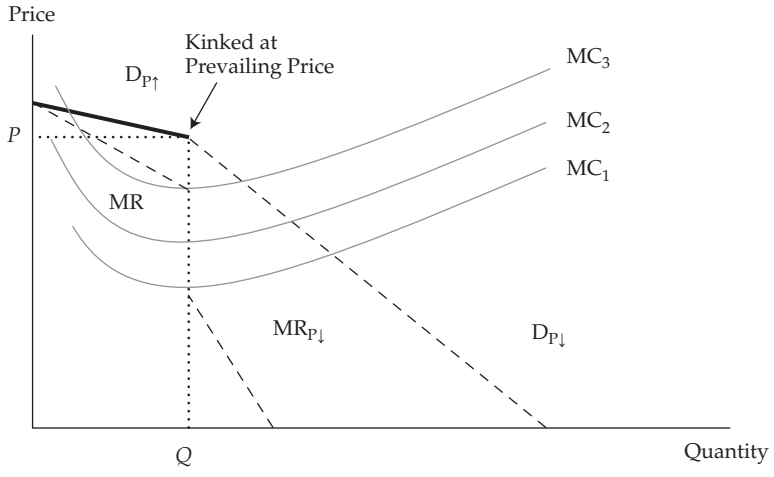
10 These examples are based on "Industry Surveys," Net Advantage Database, Standard & Poor's; and Market Share Reports, Gale Research, annual issues, as noted in McGuigan, Moyer, and Harris (2016).

Exhibit 13(A) Kinked Demand Curve in Oligopoly Market



This oligopolistic pricing strategy results in a kinked demand curve, with the two segments representing the different competitor reactions to price changes. The kink in the demand curve also yields a discontinuous marginal revenue structure, with one part associated with the price increase segment of demand and the other relating to the price decrease segment. Therefore, the firm's overall demand equals the relevant portion of $D_{P\uparrow}$ and the relevant portion of $D_{P\downarrow}$. Exhibit 13(B) represents the firm's new demand and marginal revenue schedules. The firm's demand schedule in Exhibit 13(B) is segment $D_{P\uparrow}$ and $D_{P\downarrow}$, where overall demand $D = D_{P\uparrow} + D_{P\downarrow}$.

Exhibit 13(B) Kinked Demand Curve in Oligopoly Market



Notice in Exhibit 13(B) that a wide variety of cost structures are consistent with the prevailing price. If the firm has relatively low marginal costs, MC_1, the profit-maximizing pricing rule established earlier, MR = MC, still holds for the oligopoly firm. Marginal cost can rise to MC_2 and MC_3 before the firm's profitability is challenged. If the marginal cost curve MC_2 passes through the gap in marginal revenue, the most profitable price and output combination remains unchanged at the prevailing price and original level of output.

Criticism of the kinked demand curve analysis focuses on its inability to determine what the prevailing price is from the outset. The kinked demand curve analysis does help explain why stable prices have been observed in oligopoly markets and is therefore a useful tool for analyzing such markets. However, because it cannot determine the original prevailing price, it is considered an incomplete pricing analysis.

The second pricing strategy was first developed by French economist Augustin Cournot in 1838. In the **Cournot assumption**, each firm determines its profit-maximizing production level by assuming that the other firms' output will not change. This assumption simplifies pricing strategy because there is no need to guess what the other firm will do to retaliate. It also provides a useful approach to analyzing real-world behavior in oligopoly markets. Take the most basic oligopoly market situation, a two-firm duopoly market.[11] In equilibrium, neither firm has an incentive to change output, given the other firm's production level. Each firm attempts to maximize its own profits under the assumption that the other firm will continue producing the same level of output in the future. The Cournot strategy assumes that this pattern continues until each firm reaches its long-run equilibrium position. In long-run equilibrium, output and price are stable: There is no change in price or output that will increase profits for either firm.

Consider this example of a duopoly market. Assume that the aggregate market demand has been estimated to be

$$Q_D = 450 - P$$

The supply function is represented by constant marginal cost MC = 30.

The Cournot strategy's solution can be found by setting $Q_D = q_1 + q_2$, where q_1 and q_2 represent the output levels of the two firms. Each firm seeks to maximize profit, and each firm believes the other firm will not change output as it changes its own output (Cournot's assumption). The firm will maximize profit where MR = MC. Rearranging the aggregate demand function in terms of price, we get:

$$P = 450 - Q_D = 450 - q_1 - q_2, \text{ and MC} = 30$$

Total revenue for each of the two firms is found by multiplying price and quantity:

$$\text{TR}_1 = Pq_1 = (450 - q_1 - q_2)q_1 = 450q_1 - q_1^2 - q_1q_2, \text{ and}$$

$$\text{TR}_2 = Pq_2 = (450 - q_1 - q_2)q_2 = 450q_2 - q_2q_1 - q_2^2$$

Marginal revenue is defined as the change in total revenue, given a change in sales (q_1 or q_2).[12] For the profit-maximizing output, set MR = MC, or

$$450 - 2q_1 - q_2 = 30$$

and

$$450 - q_1 - 2q_2 = 30$$

Find the simultaneous equilibrium for the two firms by solving the two equations with two unknowns:

$$450 - 2q_1 - q_2 = 450 - q_1 - 2q_2$$

---

11 The smallest possible oligopoly market is a duopoly, which is made up of only two sellers.

12 The marginal revenue formulas can be obtained using the technique introduced in Section 3.1. For the market demand function, total revenue is $P \times Q = 450Q - Q^2$ and our technique yields $\text{MR} = \Delta\text{TR}/\Delta Q = 450 - 2Q$. For the individual firms in the Cournot duopoly, $\text{MR}_1 = \Delta\text{TR}_1/\Delta q_1 = 450 - 2q_1 - q_2$, and $\text{MR}_2 = \Delta\text{TR}_2/\Delta q_2 = 450 - q_1 - 2q_2$. Each of these marginal revenue formulas is, of course, the derivative of the relevant total revenue formula with respect to the relevant quantity.

Because $q_2 = q_1$ under Cournot's assumption, insert this solution into the demand function and solve as

$$450 - 2q_1 - q_1 = 450 - 3q_1 = 30$$

Therefore, $q_1 = 140$, $q_2 = 140$, and $Q = 280$.
The price is $P = 450 - 280 = 170$.

In the Cournot strategic pricing solution, the market equilibrium price will be 170 and the aggregate output will be 280 units. This result, known as the Cournot equilibrium, differs from the perfectly competitive market equilibrium because the perfectly competitive price will be lower and the perfectly competitive output will be higher. In general, non-competitive markets have higher prices and lower levels of output in equilibrium when compared with perfect competition. In competition, the equilibrium is reached where price equals marginal cost.

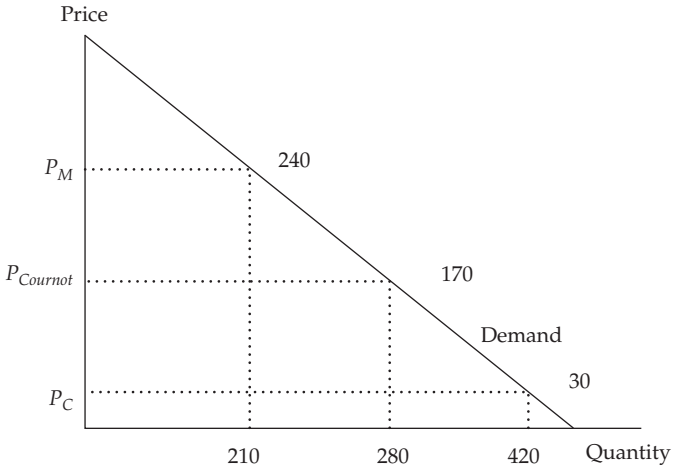
$P_C = \text{MR}_C = \text{MC}$, so $450 - Q = 30$

where P_C is the competitive firm's equilibrium price.

$Q = 420$, and $P_C = 30$.

Exhibit 14 describes the oligopoly, competitive, and monopoly market equilibrium positions, where P_M is the monopoly optimum price, P_C is the competitive price, and $P_{Cournot}$ is the oligopoly price under the Cournot assumption.

Exhibit 14 Cournot Equilibrium in Duopoly Market



In the later discussion regarding monopoly market structure, equilibrium will be established where MR = MC. That solution is also shown in Exhibit 14. The monopoly firm's demand schedule is the aggregate market demand schedule. Therefore, the solution is

$$\text{MR} = \text{MC}$$

From Footnote 10, $\text{MR} = 450 - 2Q$; therefore,

$$450 - 2Q = 30 \quad \text{and} \quad Q = 210$$

From the aggregate demand function, solve for price:

$$P_M = 450 - 210 = 240$$

Note that the Cournot solution falls between the competitive equilibrium and the monopoly solution.

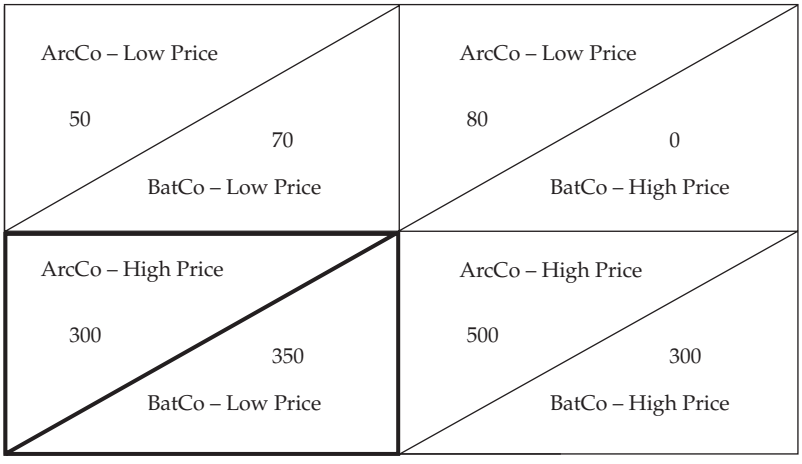
It can be shown that as the number of firms increases from two to three, from three to four, and so on, the output and price equilibrium positions move toward the competitive equilibrium solution. This result has historically been the theoretical basis for the antitrust policies established in the United States.

The third pricing strategy is attributed to one of the 1994 Nobel Prize winners, John Nash, who first developed the general concepts. In the previous analysis, the concept of market equilibrium occurs when firms are achieving their optimum remuneration under the circumstances they face. In this optimum environment, the firm has no motive to change price or output level. Existing firms are earning a normal return (zero economic profit), leaving no motive for entry to or exit from the market. All firms in the market are producing at the output level where price equals the average cost of production.

In **game theory** (the set of tools that decision makers use to consider responses by rival decision makers), the **Nash equilibrium** is present when two or more participants in a non-cooperative game have no incentive to deviate from their respective equilibrium strategies after they have considered and anticipated their opponent's rational choices or strategies. In the context of oligopoly markets, the Nash equilibrium is an equilibrium defined by the characteristic that none of the oligopolists can increase its profits by unilaterally changing its pricing strategy. The assumption is made that each participating firm does the best it can, given the reactions of its rivals. Each firm anticipates that the other firms will react to any change made by competitors by doing the best they can under the altered circumstances. The firms in the oligopoly market have interdependent actions. The actions are non-cooperative, with each firm making decisions that maximize its own profits. The firms do not collude in an effort to maximize joint profits. The equilibrium is reached when all firms are doing the best they can, given the actions of their rivals.

Exhibit 15 illustrates the duopoly result from the Nash equilibrium. Assume there are two firms in the market, ArcCo and BatCo. ArcCo and BatCo can charge high prices or low prices for the product. The market outcomes are shown in Exhibit 15.

Exhibit 15 Nash Equilibrium in Duopoly Market



For example, the top left solution indicates that when both ArcCo and BatCo offer the product at low prices, ArcCo earns a profit of 50 and BatCo earns 70. The top right solution shows that if ArcCo offers the product at a low price, BatCo earns zero profits. The solution with the maximum joint profits is the lower right equilibrium, where both firms charge high prices for the product. Joint profits are 800 in this solution.

However, the Nash equilibrium requires that each firm behaves in its own best interest. BatCo can improve its position by offering the product at low prices when ArcCo is charging high prices. In the lower left solution, BatCo maximizes its profits at 350. While ArcCo can earn 500 in its best solution, it can do so only if BatCo also agrees to charge high prices. This option is clearly not in BatCo's best interest because it can increase its return from 300 to 350 by charging lower prices.

This scenario brings up the possibility of collusion. If ArcCo agrees to share at least 51 of its 500 when both companies are charging high prices, BatCo should also be willing to charge high prices. While, in general, such collusion is unlawful in most countries, it remains a tempting alternative. Clearly, conditions in oligopolistic industries encourage collusion, with a small number of competitors and interdependent pricing behavior. Collusion is motivated by several factors: increased profits, reduced cash flow uncertainty, and improved opportunities to construct barriers to entry.

When collusive agreements are made openly and formally, the firms involved are called a **cartel**. In some cases, collusion is successful; other times, the forces of competition overpower collusive behavior. There are six major factors that affect the chances of successful collusion.[13]

1. *The number and size distribution of sellers.* Successful collusion is more likely if the number of firms is small or if one firm is dominant. Collusion becomes more difficult as the number of firms increases or if the few firms have similar market shares. When the firms have similar market shares, the competitive forces tend to overshadow the benefits of collusion.
2. *The similarity of the products.* When the products are homogeneous, collusion is more successful. The more differentiated the products, the less likely it is that collusion will succeed.
3. *Cost structure.* The more similar the firms' cost structures, the more likely it is that collusion will succeed.
4. *Order size and frequency.* Successful collusion is more likely when orders are frequent, received on a regular basis, and relatively small. Frequent small orders, received regularly, diminish the opportunities and rewards for cheating on the collusive agreement.
5. *The strength and severity of retaliation.* Oligopolists will be less likely to break the collusive agreement if the threat of retaliation by the other firms in the market is severe.
6. *The degree of external competition.* The main reason to enter into the formal collusion is to increase overall profitability of the market, and rising profits attract competition. For example, in 2016 the average extraction cost of a barrel of crude oil from Saudi Arabia was approximately $9, while the average cost from United States shale oil fields was roughly $23.50. The cost of extracting oil from the Canadian tar sands in 2016 was roughly $27 per barrel. It is more likely that crude oil producers in the gulf countries will successfully collude because of the similarity in their cost structures (roughly $9–$10 per barrel). If OPEC had held crude oil prices down below $30 per barrel, there would have been a viable economic argument to develop US shale oil fields through fracking or expand extraction from Canada's tar sands. OPEC's successful cartel raised crude oil prices to the point where outside sources became economically possible and in doing so increased the competition the cartel faces.[14]

13 McGuigan, Moyer, and Harris (2016).
14 "Barrel Breakdown," *Wall Street Journal*, April 15, 2016.

There are other possible oligopoly strategies that are associated with decision making based on game theory. The Cournot equilibrium and the Nash equilibrium are examples of specific strategic games. A strategic game is any interdependent behavioral choice employed by individuals or groups that share a common goal (e.g., military units, sports teams, or business decision makers). Another prominent decision-making strategy in oligopolistic markets is the first-mover advantage in the **Stackelberg model**, named after the economist who first conceptualized the strategy.[15] The important difference between the Cournot model and the Stackelberg model is that Cournot assumes that in a duopoly market, decision making is simultaneous, while Stackelberg assumes that decisions are made sequentially. In the Stackelberg model, the leader firm chooses its output first and then the follower firm chooses after observing the leader's output. It can be shown that the leader firm has a distinct advantage, being a first mover.[16] In the Stackelberg game, the leader can aggressively overproduce to force the follower to scale back its production or even punish or eliminate the weaker opponent. This approach is sometimes referred to as a "top dog" strategy.[17] The leader earns more than in Cournot's simultaneous game, while the follower earns less. Many other strategic games are possible in oligopoly markets. The important conclusion is that the optimal strategy of the firm depends on what its adversary does. The price and marginal revenue the firm receives for its product depend on both its decisions and its adversary's decisions.

5.2 Supply Analysis in Oligopoly Markets

As in monopolistic competition, the oligopolist does not have a well-defined supply function. That is, there is no way to determine the oligopolist's optimal levels of output and price independent of demand conditions and competitor's strategies. However, the oligopolist still has a cost function that determines the optimal level of supply. Therefore, the profit-maximizing rule established earlier is still valid: The level of output that maximizes profit is where MR = MC. The price to charge is determined by what price consumers are willing to pay for that quantity of the product. Therefore, the equilibrium price comes from the demand curve, while the output level comes from the relationship between marginal revenue and marginal cost.

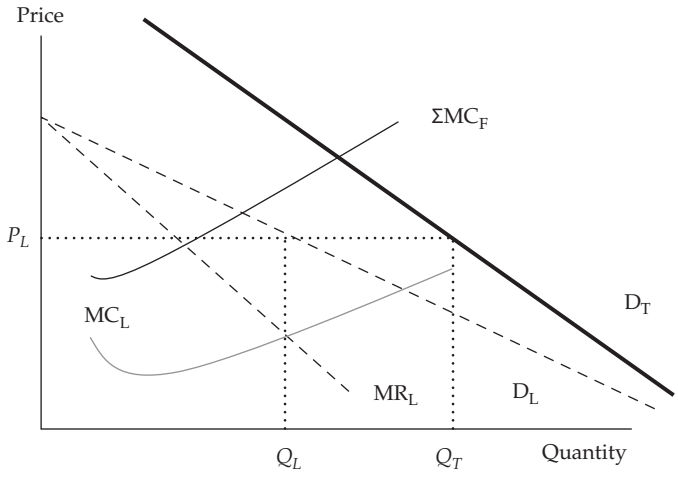
Consider an oligopoly market in which one of the firms is dominant and thus able to be the price leader. Dominant firms generally have 40 percent or greater market share. When one firm dominates an oligopoly market, it does so because it has greater capacity, has a lower cost structure, was first to market, or has greater customer loyalty than other firms in the market.

Assuming there is no collusion, the dominant firm becomes the price maker, and therefore its actions are similar to monopoly behavior in its segment of the market. The other firms in the market follow the pricing patterns of the dominant firm. Why wouldn't the price followers attempt to gain market share by undercutting the dominant firm's price? The most common explanation is that the dominant firm's supremacy often stems from a lower cost of production. Usually, the price followers would rather charge a price that is even higher than the dominant firm's price choice. If they attempt to undercut the dominant firm, the followers risk a price war with a lower-cost producer that can threaten their survival. Some believe that one explanation for the price leadership position of the dominant firm is simply convenience. Only one firm has to make the pricing decisions, and the others can simply follow its lead.

15 Von Stackelberg (1952). See also Kelly (2011) for a comparison between the Cournot and Stackelberg equilibriums.
16 Nicholson and Snyder (2016).
17 Fudenberg and Tirole (1984).

Exhibit 16 establishes the dominant firm's pricing decision. The dominant firm's demand schedule, D_L, is a substantial share of the total market demand, D_T. The low-cost position of the dominant firm is represented by its marginal cost, MC_L. The sum of the marginal costs of the price followers is established as ΣMC_F and represents a higher cost of production than that of the price leader.

Exhibit 16 Dominant Oligopolist's Price Leadership



There is an important reason why the total demand curve and the leader demand curve are not parallel in Exhibit 16: Remember that the leader is the low-cost producer. Therefore, as price decreases, fewer of the smaller suppliers will be able to profitably remain in the market, and several will exit because they do not want to sell below cost. Therefore, the leader will have a larger market share as P decreases, which implies that Q_L increases at a low price, exactly as shown by a steeper D_T in the diagram.

The price leader identifies its profit-maximizing output where $MR_L = MC_L$, at output Q_L. This is the quantity it wants to supply; however, the price it will charge is determined by its segment of the total demand function, D_L. At price P_L, the dominant firm will supply quantity Q_L of total demand, D_T. The price followers will supply the difference to the market, $(Q_T - Q_L) = Q_F$. Therefore, neither the dominant firm nor the follower firms have a single functional relationship that determines the quantity supplied at various prices.

5.3 Optimal Price and Output in Oligopoly Markets

From the discussion above, clearly there is no single optimum price and output analysis that fits all oligopoly market situations. The interdependence among the few firms that make up the oligopoly market provides a complex set of pricing alternatives, depending on the circumstances in each market. In the case of the kinked demand curve, the optimum price is the prevailing price at the kink in the demand function. However, as previously noted, the kinked demand curve analysis does not provide insight into what established the prevailing price in the first place.

Perhaps the case of the dominant firm, with the other firms following the price leader, is the most obvious. In that case, the optimal price is determined at the output level where MR = MC. The profit-maximizing price is then determined by the output position of the segment of the demand function faced by the dominant firm. The price

followers have little incentive to change the leader's price. In the case of the Cournot assumption, each firm assumes that the other firms will not alter their output following the dominant firm's selection of its price and output level.

Therefore, again, the optimum price is determined by the output level where MR = MC. In the case of the Nash equilibrium, each firm will react to the circumstances it faces, maximizing its own profit. These adjustments continue until there are stable prices and levels of output. Because of the interdependence, there is no certainty as to the individual firm's price and output level.

5.4 Factors Affecting Long-Run Equilibrium in Oligopoly Markets

Long-run economic profits are possible for firms operating in oligopoly markets. However, history has shown that, over time, the market share of the dominant firm declines. Profits attract entry by other firms into the oligopoly market. Over time, the marginal costs of the entrant firms decrease because they adopt more efficient production techniques, the dominant firm's demand and marginal revenue shrink, and the profitability of the dominant firm declines. In the early 1900s, J.P. Morgan, Elbert Gary, Andrew Carnegie, and Charles M. Schwab created the United States Steel Corporation (US Steel). When it was first formed in 1901, US Steel controlled 66 percent of the market. By 1920, US Steel's market share had declined to 46 percent, and by 1925 its market share was 42 percent.

In the long run, optimal pricing strategy must include the reactions of rival firms. History has proven that pricing wars should be avoided because any gains in market share are temporary. Decreasing prices to drive competitors away lowers total revenue to all participants in the oligopoly market. Innovation may be a way—though sometimes an uneconomical one—to maintain market leadership.

OLIGOPOLIES: APPEARANCE VERSUS BEHAVIOR

When is an oligopoly not an oligopoly? There are two extreme cases of this situation. A normal oligopoly has a few firms producing a differentiated good, and this differentiation gives them pricing power.

At one end of the spectrum, we have the oligopoly with a credible threat of entry. In practice, if the oligopolists are producing a good or service that can be easily replicated, has limited economies of scale, and is not protected by brand recognition or patents, they will not be able to charge high prices. The easier it is for a new supplier to enter the market, the lower the margins. In practice, this oligopoly will behave very much like a perfectly competitive market.

At the opposite end of the spectrum, we have the case of the cartel. Here, the oligopolists collude and act as if they were a single firm. In practice, a very effective cartel enacts a cooperative strategy. As shown in Section 5.1, instead of going to a Nash equilibrium, the cartel participants go to the more lucrative (for them) cooperative equilibrium.

A cartel may be explicit (that is, based on a contract) or implicit (based on signals). An example of signals in a duopoly would be that one of the firms reduces its prices and the other does not. Because the firm not cutting prices refuses to start a price war, the firm that cut prices may interpret this signal as a "suggestion" to raise prices to a higher level than before, so that profits may increase for both.

MONOPOLY

6

Monopoly market structure is at the opposite end of the spectrum from perfect competition. For various reasons, there are significant barriers to entry such that a single firm produces a highly specialized product and faces no threat of competition. There are no good substitutes for the product in the relevant market, and the market demand function is the same as the individual firm's demand schedule. *The distinguishing characteristics of monopoly are that a single firm represents the market and significant barriers to entry exist*. Exhibit 1 identified the characteristics of monopoly markets:

1. There is a single seller of a highly differentiated product.
2. The product offered by the seller has no close substitute.
3. Entry into the market is very difficult, with high costs and significant barriers to competition.
4. The firm has considerable pricing power.
5. The product is differentiated through non-price strategies such as advertising.

Monopoly markets are unusual. With a single seller dominating the market, power over price decisions is significant. For a single seller to achieve this power, there must be factors that allow the monopoly to exist. One obvious source of monopoly power would be a patent or copyright that prevents other firms from entering the market. Patent and copyright laws exist to reward intellectual capital and investment in research and development. In so doing, they provide significant barriers to entry.

Another possible source of market power is control over critical resources used for production. One example is De Beers Consolidated Mines Limited. De Beers owned or controlled all diamond mining operations in South Africa and established pricing agreements with other important diamond producers. In doing so, De Beers was able to control the prices for cut diamonds for decades. Technically, De Beers was a near-monopoly dominant firm rather than a pure monopoly, although its pricing procedure for cut diamonds resembled monopoly behavior.

Perhaps the most common form of monopolistic market power occurs as the result of government-controlled authorization. In most urban areas, a single source of water and sewer services is offered. In some cases, these services are offered by a government-controlled entity. In other cases, private companies provide the services under government regulation. Such "natural" monopolies require a large initial investment that benefits from economies of scale; therefore, government may authorize a single seller to provide a certain service because having multiple sellers would be too costly. For example, electricity in most markets is provided by a single seller. Economies of scale result when significant capital investment benefits from declining long-run average costs. In the case of electricity, a large gas-fueled power plant producing electricity for a large area is substantially more efficient than having a small diesel generator for every building. That is, the average cost of generating and delivering a kilowatt of electricity will be substantially lower with the single power station, but the initial fixed cost of building the power station and the lines delivering electricity to each home, factory, and office will be very high.

In the case of natural monopolies, limiting the market to a single seller is considered beneficial to society. One water and sewer system is deemed better for the community than dozens of competitors because building multiple infrastructures for running water and sewer service would be particularly expensive and complicated. One electrical power grid supplying electricity for a community can make large capital investments in generating plants and lower the long-run average cost, while multiple power grids would lead to a potentially dangerous maze of wires. Clearly, not all monopolies are in a position to make significant economic profits. Regulators, such as public utility

commissions in the United States, attempt to determine what a normal return for the monopoly owners' investment should be, and they set prices accordingly. Nevertheless, monopolists attempt to maximize profits.

Not all monopolies originate from "natural" barriers. For some monopolists, barriers to entry do not derive from increasing returns to scale. We mentioned that marketing and brand loyalty are sources of product differentiation in monopolistic competition. In some highly successful cases, strong brand loyalty can become a formidable barrier to entry. For example, if the Swiss watchmaker Rolex is unusually successful in establishing brand loyalty, so that its customers think there is no close substitute for its product, then the company will have monopoly-like pricing power over its market.

The final potential source of market power is the increasing returns associated with network effects. Network effects result from synergies related to increasing market penetration. By achieving a critical level of adoption, Microsoft was able to extend its market power through the network effect—for example, because most computer users know how to use Microsoft Word. Therefore, for firms, Word is cheaper to adopt than other programs because almost every new hire will be proficient in using the software and will need no further training. At some level of market share, a network-based product or service (think of Facebook or eBay) reaches a point where each additional share point increases the probability that another user will adopt.[18] These network effects increase the value to other potential adopters. In Microsoft's case, the network effects crowded out other potential competitors, including Netscape's internet browser, that might have led to applications bypassing Windows. Eventually, Microsoft's operating system's market share reached 92 percent of the global market. Similar situations occur in financial markets: If a publicly listed share or a derivative contract is more frequently traded on a certain exchange, market participants wishing to sell or buy the security will go to the more liquid exchange because they expect to find a better price and faster execution there.

6.1 Demand Analysis in Monopoly Markets

The monopolist's demand schedule is the aggregate demand for the product in the relevant market. Because of the income effect and the substitution effect, demand is negatively related to price, as usual. The slope of the demand curve is negative and therefore downward sloping. The general form of the demand relationship is

$$Q_D = a - bP \quad \text{or, rewritten,} \quad P = a/b - (1/b)Q_D$$

$$\text{Therefore, total revenue} = \text{TR} = P \times Q = (a/b)Q_D - (1/b)Q_D^2$$

Marginal revenue is the change in revenue given a change in the quantity demanded. Because an increase in quantity requires a lower price, the marginal revenue schedule is steeper than the demand schedule. If the demand schedule is linear, then the marginal revenue curve is twice as steep as the demand schedule.[19]

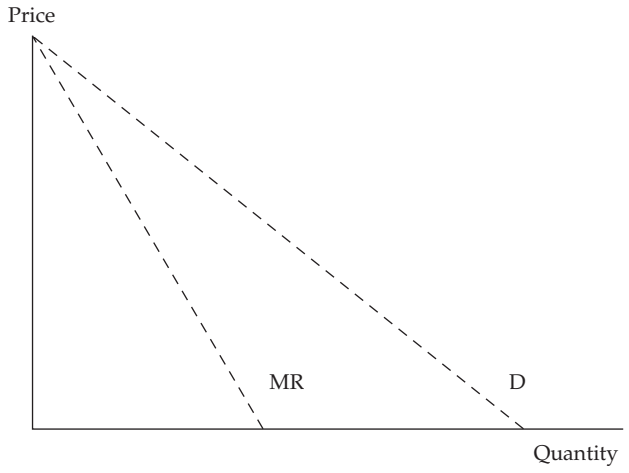
$$\text{MR} = \Delta\text{TR}/\Delta Q = (a/b) - (2/b)Q_D$$

The demand and marginal revenue relationship is expressed in Exhibit 17.

18 When a network-based device reaches a 30 percent share, the next 50 percentage points are cheaper to promote, according to McGuigan, Moyer, and Harris (2016).

19 Marginal revenue can be found using the technique shown in Section 3.1 or, for readers who are familiar with calculus, by taking the derivative of the total revenue function: $\text{MR} = \Delta\text{TR}/\Delta Q = (a/b) - (2/b)Q_D$.

Exhibit 17 Monopolist's Demand and Marginal Revenue



Suppose a company operating on a remote island is the single seller of natural gas. Demand for its product can be expressed as

$Q_D = 400 - 0.5P$, which can be rearranged as
$P = 800 - 2Q_D$

Total revenue is $P \times Q = \text{TR} = 800Q_D - 2Q_D^2$, and marginal revenue is MR = 800 $- 4Q_D$.[20]

In Exhibit 17, the demand curve's intercept is 800 and the slope is –2. The marginal revenue curve in Exhibit 17 has an intercept of 800 and a slope of –4.

Average revenue is TR/Q_D; therefore, AR = $800 - 2Q_D$, which is the same as the demand function. In the monopoly market model, average revenue is the same as the market demand schedule.

6.2 Supply Analysis in Monopoly Markets

A monopolist's supply analysis is based on the firm's cost structure. As in the market structures of monopolistic competition and oligopoly, the monopolist does not have a well-defined supply function that determines the optimal output level and the price to charge. The optimal output is the profit-maximizing output level. The profit-maximizing level of output occurs where marginal revenue equals marginal cost, MR = MC.

Assume the natural gas company has determined that its total cost can be expressed as

$$\text{TC} = 20{,}000 + 50Q + 3Q^2$$

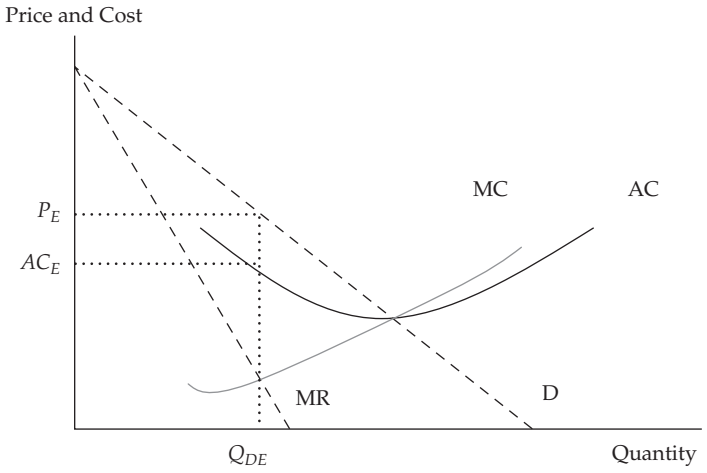
Marginal cost is $\Delta\text{TC}/\Delta Q = \text{MC} = 50 + 6Q$.[21]

Supply and demand can be combined to determine the profit-maximizing level of output. Exhibit 18 combines the monopolist's demand and cost functions.

20 MR = $\Delta\text{TR}/\Delta Q = 800 - 4Q$; see footnote 16.

21 The marginal cost equation can be found in this case by applying the technique used to find the marginal revenue equation in Section 3.1, or by taking the derivative of the total cost function.

Exhibit 18 Monopolist's Demand, Marginal Revenue, and Cost Structures



In Exhibit 18, the demand and marginal revenue functions are clearly defined by the aggregate market. However, the monopolist does not have a supply curve. The quantity that maximizes profit is determined by the intersection of MC and MR, Q_{DE}.

The price consumers are willing to pay for this level of output is P_E, as determined by the demand curve, P_E.

The profit-maximizing level of output is MR = MC: $800 - 4Q_D = 50 + 6Q_D$; therefore, $Q_D = 75$ when profit is maximized.

Total profit equals total revenue minus total cost:

$$\pi = 800Q - 2Q_D^2 - \left(20{,}000 + 50Q_D + 3Q_D^2\right) = -20{,}000 + 750Q_D - 5Q_D^2$$

Profit is represented by the difference between the area of the rectangle $Q_{DE} \times P_E$, representing total revenue, and the area of the rectangle $Q_{DE} \times AC_E$, representing total cost.

MONOPOLISTS AND THEIR INCENTIVES

In theoretical models, which usually take product quality and technology as given, monopolists can choose to vary either price or quantity. In real life, they also can vary their product.

A monopolist can choose to limit quality if producing a higher-quality product is costly and higher quality does not increase profits accordingly. For example, the quality of domestically produced cars in most developed countries improved dramatically once imports became more available. Before the opening of borders to imports, the single incumbent that dominated the market (for example, Fiat in Italy) or the small group of incumbents acting as a collusive oligopoly (such as the Detroit "Big Three" in the United States) were the effective monopolists of their domestic automobile markets. Rust corrosion, limited reliability, and poor gas mileage were common.[22]

Similarly, regulated utilities may have limited incentives to innovate. Several studies, including Gómez-Ibáñez (2003), have found that state-owned and other monopoly telephone utilities tended to provide very poor service before competition was introduced. Poor service may not be limited to poor connection quality but may also include extensive delays in adding new users and limited introduction of new services, such as caller ID or automatic answering services.

22 For more on this topic, see Banker, Khosla, and Sinha (1998).

Intuitively, a monopolist will not spend resources on quality control, research and development, or even customer relations unless there is a threat of entry of a new competitor or unless there is a clear link between such expenses and a profit increase. In contrast, in competitive markets, including oligopoly, innovation and quality are often ways to differentiate the product and increase profits.

6.3 Optimal Price and Output in Monopoly Markets

Continuing the natural gas example from above, the total profit function can be solved using the quadratic formula.[23] Another method to solve the profit function is to evaluate $\Delta\pi/\Delta Q_D$ and set it equal to zero. This identifies the point at which profit is unaffected by changes in output.[24] Of course, this will give the same result as we found by equating marginal revenue with marginal cost. The monopoly will maximize profits when $Q^* = 75$ units of output and the price is set from the demand curve at 650.

$P^* = 800 - 2(75) = 650$ per unit

To find total maximum profits, substitute these values into the profit function above:

$$\pi = -20{,}000 + 750Q_D - 5Q_D^2 = -20{,}000 + 750(75) - 5\left(75^2\right) = 8{,}125$$

Note that the price and output combination that maximizes profit occurs in the elastic portion of the demand curve in Exhibit 18. This must be so because marginal revenue and marginal cost will always intersect where marginal revenue is positive. This fact implies that quantity demanded responds more than proportionately to prices changes, i.e. demand is elastic, at the point at which MC = MR. As noted earlier, the relationship between marginal revenue and price elasticity, E_P, is:

$MR = P[1 - 1/E_P]$

In monopoly, MR = MC; therefore,

$P[1 - 1/E_P] = MC$

The firm can use this relationship to determine the profit-maximizing price if the firm knows its cost structure and the price elasticity of demand, E_P. For example, assume the firm knows that its marginal cost is constant at 75 and recent market analysis indicates that price elasticity is estimated to be 1.5. The optimal price is solved as

$P[1 - 1/1.5] = 75$ and

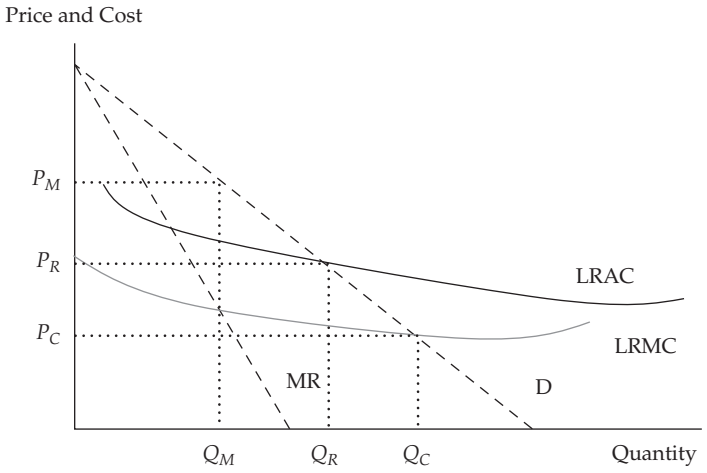
$P = 225$

Exhibit 18 indicated that the monopolist wants to produce at Q_E and charge the price of P_E. Suppose this is a natural monopoly that is operating as a government franchise under regulation. Natural monopolies are usually found where production is based on significant economies of scale and declining cost structure in the market. Examples include electric power generation, natural gas distribution, and the water and sewer industries. These are often called public utilities. Exhibit 19 illustrates such a market in long-run equilibrium.

23 The quadratic formula, where $aQ^2 + bQ + c = 0$, is $Q = \left\{-b \pm \sqrt{\left(b^2 - 4ac\right)}\right\} \Big/ 2a$.

24 Maximum profit occurs where $\Delta\pi/\Delta Q_D = 0 = 750 - 10Q_D$. Therefore, profits are maximized at $Q_D = 75$.

Exhibit 19 Natural Monopoly in a Regulated Pricing Environment



In Exhibit 19, three possible pricing and output solutions are presented. The first is what the monopolist would do without regulation: The monopolist would seek to maximize profits by producing Q_M units of the product, where long-run marginal cost equals marginal revenue, LRMC = MR. To maximize profits, the monopolist would raise the price to the level the demand curve will accept, P_M.

In perfect competition, the price and output equilibrium occurs where price is equal to the marginal cost of producing the incremental unit of the product. In a competitive market, the quantity produced would be higher, Q_C, and the price lower, P_C. For this regulated monopoly, the competitive solution would be unfair because at output Q_C, the price P_C would not cover the average cost of production. One possibility is to subsidize the difference between the long-run average cost, LRAC, and the competitive price, P_C, for each unit sold.

Another solution is for the regulator to set the price at the point where long-run average cost equals average revenue. Recall that the demand curve represents the average revenue the firm receives at each output level. The government regulator will attempt to determine the monopolistic firm's long-run average cost and set the output and price so that the firm receives a fair return on the owners' invested capital. The regulatory solution is output level Q_R, with the price set at P_R. Therefore, the regulatory solution is found between the unregulated monopoly equilibrium and the competitive equilibrium.

6.4 Price Discrimination and Consumer Surplus

Monopolists can either be more or less effective in taking advantage of their market structure. At one extreme, we have a monopolist that charges prices and supplies quantities that are the same as they would be in perfect competition; this scenario may be a result of regulation or threat of entry (if the monopolist charged more, another company could come in and price the former monopolist out of the market). At the opposite extreme, hated by all consumers and economists, is the monopolist that extracts the entire consumer surplus. This scenario is called **first-degree price discrimination**, where a monopolist can charge each customer the highest price the customer is willing to pay. This is called price discrimination because the monopolist charges a different price to each client. How can this be? For example, if the monopolist knows the exact demand schedule of the customer, then the monopolist can capture the entire consumer surplus. In practice, the monopolist can measure how often the product is used and charges the customer the highest price the consumer is

willing to pay for that unit of good. Another possibility is that public price disclosure is non-existent, so that no customer knows what the other customers are paying. Interestingly, not every consumer is worse off in this case, because some consumers may be charged a price that is below that of perfect competition, as long as the marginal revenue exceeds the marginal cost.

In **second-degree price discrimination** the monopolist offers a menu of quantity-based pricing options designed to induce customers to self-select based on how highly they value the product. Such mechanisms include volume discounts, volume surcharges, coupons, product bundling, and restrictions on use. In practice, producers can use not just the quantity but also the quality (e.g., "professional grade") to charge more to customers that value the product highly.

Third-degree price discrimination happens when customers are segregated by demographic or other traits. For example, some econometric software is licensed this way: A student version can handle only small datasets and is sold for a low price; a professional version can handle very large datasets and is sold at a much higher price because corporations need to compute the estimates for their business and are therefore willing to pay more for a license. Another example is that airlines know that passengers who want to fly somewhere and come back the same day are most likely business people; therefore, one-day roundtrip tickets are generally more expensive than tickets with a return flight at a later date or over a weekend.

Price discrimination has many practical applications when the seller has pricing power. The best way to understand how this concept works is to think of consumer surplus: As seen in this reading, a consumer may be willing to pay more for the first unit of a good, but to buy a second unit she will want to pay a lower price, thus getting a better deal on the first unit. In practice, sellers can sometimes use income and substitution effects to their advantage. Think of something you often buy, perhaps lunch at your favorite café. How much would you be willing to pay for a "lunch club membership card" that would allow you to purchase lunches at, say, half price? If the café could extract from you the maximum amount each month that you would be willing to pay for the half-price option, then it would successfully have removed the income effect from you in the form of a monthly fixed fee. Notice that a downward-sloping demand curve implies that you would end up buying more lunches each month than before you purchased the discount card, even though you would be no better or worse off than before. This is a way that sellers are sometimes able to extract consumer surplus by means of creative pricing schemes. It's a common practice among big-box retailers, sports clubs, and other users of what is called "two-part tariff pricing," as in the example below.

EXAMPLE 3

Price Discrimination

Nicole's monthly demand for visits to her health club is given by the following equation: $Q_D = 20 - 4P$, where Q_D is visits per month and P is euros per visit. The health club's marginal cost is fixed at €2 per visit.

1. Draw Nicole's demand curve for health club visits per month.
2. If the club charged a price per visit equal to its marginal cost, how many visits would Nicole make per month?
3. How much consumer surplus would Nicole enjoy at that price?
4. How much could the club charge Nicole each month for a membership fee?

Solution to 1:

$Q_D = 20 - 4P$, so when $P = 0$, $Q_D = 20$. Inverting, $P = 5 - 0.25Q_D$, so when $Q = 0$, $P = 5$.

Solution to 2:

$Q_D = 20 - 4(2) = 12$. Nicole would make 12 visits per month at a price of €2 per visit.

Solution to 3:

Nicole's consumer surplus can be measured as the area under her demand curve and above the price she pays for a total of 12 visits, or (0.5)(12)(3) = 18. Nicole would enjoy a consumer surplus of €18 per month.

Solution to 4:

The club could extract all of Nicole's consumer surplus by charging her a monthly membership fee of €18 plus a per-visit price of €2. This pricing method is called a two-part tariff because it assesses one price per unit of the item purchased plus a per-month fee (sometimes called an "entry fee") equal to the buyer's consumer surplus evaluated at the per-unit price.

6.5 Factors Affecting Long-Run Equilibrium in Monopoly Markets

The unregulated monopoly market structure can produce economic profits in the long run. In the long run, all factors of production are variable, while in the short run, some factors of production are fixed. Generally, the short-run factor that is fixed is the capital investment, such as the factory, machinery, production technology, available arable land, and so forth. The long-run solution allows for all inputs, including technology, to change. In order to maintain a monopoly market position in the long run, the firm must be protected by substantial and ongoing barriers to entry. If the monopoly position is the result of a patent, then new patents must be continuously added to prevent the entry of other firms into the market.

For regulated monopolies, such as natural monopolies, there are a variety of long-run solutions. One solution is to set the price equal to marginal cost, P = MC. However, that price will not likely be high enough to cover the average cost of production, as Exhibit 19 illustrated. The answer is to provide a subsidy sufficient to compensate the firm. The national rail system in the United States, Amtrak, is an example of a regulated monopoly operating with a government subsidy.

National ownership of the monopoly is another solution. Nationalization of the natural monopoly has been a popular solution in Europe and other parts of the world. The United States has generally avoided this potential solution. One problem with this arrangement is that once a price is established, consumers are unwilling to accept price increases, even as factor costs increase. Politically, raising prices on products from government-owned enterprises is highly unpopular.

Establishing a governmental entity that regulates an authorized monopoly is another popular solution. Exhibit 19 illustrated the appropriate decision rule. The regulator sets price equal to long-run average cost, P_R = LRAC. This solution assures that investors will receive a normal return for the risk they are taking in the market. Given that no other competitors are allowed, the risk is lower than in a highly competitive market environment. The challenge facing the regulator is determining the authentic risk-related return and the monopolist's realistic long-run average cost.

The final solution is to franchise the monopolistic firm through a bidding war. Again, the public goal is to select the winning firm based on price equaling long-run average cost. Retail outlets at rail stations and airports and concession outlets at stadiums are examples of government franchises. The long-run success of the monopoly franchise depends on its ability to meet the goal of pricing its products at the level of its long-run average cost.

EXAMPLE 4

Monopolies and Efficiency

Are monopolies *always* inefficient?

A No, because if they charge more than average cost they are nationalized.

B Yes, because they charge all consumers more than perfectly competitive markets would.

C No, because economies of scale and regulation (or threat of entry) may give a better outcome for buyers than perfect competition.

Solution:

C is correct. Economies of scale and regulation may make monopolies more efficient than perfect competition.

IDENTIFICATION OF MARKET STRUCTURE 7

Monopoly markets and other situations where companies have pricing power can be inefficient because producers constrain output to cause an increase in prices. Therefore, there will be less of the good being consumed and it will be sold at a higher price, which is generally inefficient for the overall market. As a result, many countries have introduced competition law to regulate the degree of competition in many industries.

Market power in the real world is not always as clear as it is in textbook examples. Governments and regulators often have the difficult task of measuring market power and establishing whether a firm has a dominant position that may resemble a monopoly. A few historical examples of this are as follows:

1. In the 1990s, US regulators prosecuted agricultural corporation Archer Daniels Midland for conspiring with Japanese competitors to fix the price of lysine, an amino acid used as an animal feed additive. The antitrust action resulted in a settlement that involved over US$100 million in fines paid by the cartel members.
2. In the 1970s, US antitrust authorities broke up the local telephone monopoly, leaving AT&T the long-distance business (and opening that business to competitors), and required AT&T to divest itself of the local telephone companies it owned. This antitrust decision brought competition, innovation, and lower prices to the US telephone market.
3. European regulators (specifically, the European Commission) have affected the mergers and monopoly positions of European corporations (as in the case of the companies Roche, Rhone-Poulenc, and BASF, which were at the center of a vitamin price-fixing case) as well as non-European companies (such as Intel) that

do business in Europe. Moreover, the merger between the US company General Electric and the European company Honeywell was denied by the European Commission on grounds of excessive market concentration.

Quantifying excessive market concentration is difficult. Sometimes, regulators need to measure whether something that has not yet occurred might generate excessive market power. For example, a merger between two companies might allow the combined company to be a monopolist or quasi monopolist in a certain market.

A financial analyst hearing news about a possible merger should always consider the impact of competition law (sometimes called antitrust law)—that is, whether a proposed merger may be blocked by regulators in the interest of preserving a competitive market.

7.1 Econometric Approaches

How should one measure market power? The theoretical answer is to estimate the elasticity of demand and supply in a market. If demand is very elastic, the market must be very close to perfect competition. If demand is rigid (inelastic), companies *may* have market power. This is the approach taken in the cellophane case mentioned in Section 3.1.2.

From the econometric point of view, this estimation requires some attention. The problem is that observed price and quantity are the equilibrium values of price and quantity and do not represent the value of either supply or demand. Technically, this is called the problem of endogeneity, in the sense that the equilibrium price and quantity are jointly determined by the interaction of demand and supply. Therefore, to have an appropriate estimation of demand and supply, we will need to use a model with two equations, namely, an equation of demanded quantity (as a function of price, income of the buyers, and other variables) and an equation of supplied quantity (as a function of price, production costs, and other variables). The estimated parameters will then allow us to compute elasticity.

Regression analysis is useful in computing elasticity but requires a large number of observations. Therefore, one may use a time-series approach and, for example, look at 20 years of quarterly sales data for a market. However, the market structure may have changed radically over those 20 years, and the estimated elasticity may not apply to the current situation. Moreover, the supply curve may change due to a merger among large competitors, and the estimation based on past data may not be informative regarding the future state of the market post merger.

An alternative approach is a cross-sectional regression analysis. Instead of looking at total sales and average prices in a market over time (the time-series approach mentioned above), we can look at sales from different companies in the market during the same year, or even at single transactions from many buyers and companies. Clearly, this approach requires substantial data-gathering effort, and therefore, this estimation method can be complicated. Moreover, different specifications of the explanatory variables (for example, using total GDP rather than median household income or per-capita GDP to represent income) may sometimes lead to dramatically different estimates.

7.2 Simpler Measures

Trying to avoid the above drawbacks, analysts often use simpler measures to estimate elasticity. The simplest measure is the concentration ratio, which is the sum of the market shares of the largest N firms. To compute this ratio, one would, for example, add the sales values of the largest 10 firms and divide this figure by total market sales. This number is always between zero (perfect competition) and 100 percent (monopoly).

The main advantage of the concentration ratio is that it is simple to compute, as shown above. The disadvantage is that it does not directly quantify market power. In other words, is a high concentration ratio a clear signal of monopoly power? The analysis of entry in Section 2 explains clearly that this is not the case: A company may be the only incumbent in a market, but if the barriers to entry are low, the simple presence of a *potential* entrant may be sufficient to convince the incumbent to behave like a firm in perfect competition. For example, a sugar wholesaler may be the only one in a country, but the knowledge that other large wholesalers in the food industry might easily add imported sugar to their range of products should convince the sugar wholesaler to price its product as if it were in perfect competition.

Another disadvantage of the concentration ratio is that it tends to be unaffected by mergers among the top market incumbents. For example, if the largest and second-largest incumbents merge, the pricing power of the combined entity is likely to be larger than that of the two pre-existing companies. But the concentration ratio may not change much.

CALCULATING THE CONCENTRATION RATIO

Suppose there are eight producers of a certain good in a market. The largest producer has 35 percent of the market, the second largest has 25 percent, the third has 20 percent, the fourth has 10 percent, and the remaining four have 2.5 percent each. If we computed the concentration ratio of the top three producers, it would be 35 + 25 + 20 = 80 percent, while the concentration ratio of the top four producers would be 35 + 25 + 20 + 10 = 90 percent.

If the two largest companies merged, the new concentration ratio for the top three producers would be 60 (the sum of the market shares of the merged companies) + 20 + 10 = 90 percent, and the concentration ratio for the four top producers would be 92.5 percent. Therefore, this merger affects the concentration ratio very mildly, even though it creates a substantial entity that controls 60 percent of the market.

For example, the effect of consolidation in the US retail gasoline market has resulted in increasing degrees of concentration. In 1992, the top four companies in the US retail gasoline market shared 33 percent of the market. By 2001, the top four companies controlled 78 percent of the market (Exxon Mobil 24 percent, Shell 20 percent, BP/Amoco/Arco 18 percent, and Chevron/Texaco 16 percent).

To avoid the known issues with concentration ratios, economists O.C. Herfindahl and A.O. Hirschman suggested an index where the market shares of the top N companies are first squared and then added. If one firm controls the whole market (a monopoly), the Herfindahl–Hirschman index (HHI) equals 1. If there are M firms in the industry with equal market shares, then the HHI equals $(1/M)$. This provides a useful gauge for interpreting an HHI. For example, an HHI of 0.20 would be analogous to having the market shared equally by 5 firms.

The HHI for the top three companies in the example in the box above would be $0.35^2 + 0.25^2 + 0.20^2 = 0.225$ before the merger, while after the merger, it would be $0.60^2 + 0.20^2 + 0.10^2 = 0.410$, which is substantially higher than the initial 0.225. The HHI is widely used by competition regulators; however, just like the concentration ratio, the HHI does not take the possibility of entry into account, nor does it consider the elasticity of demand. Therefore, the HHI has limited use for a financial analyst trying to estimate the potential profitability of a company or group of companies.

EXAMPLE 5

The Herfindahl–Hirschman Index

Suppose a market has 10 suppliers, each of them with 10 percent of the market. What are the concentration ratio and the HHI of the top four firms?

A Concentration ratio 4 percent and HHI 40

B Concentration ratio 40 percent and HHI 0.4

C Concentration ratio 40 percent and HHI 0.04

Solution:

C is correct. The concentration ratio for the top four firms is 10 + 10 + 10 + 10 = 40 percent, and the HHI is $0.10^2 \times 4 = 0.01 \times 4 = 0.04$.

SUMMARY

In this reading, we have surveyed how economists classify market structures. We have analyzed the distinctions between the different structures that are important for understanding demand and supply relations, optimal price and output, and the factors affecting long-run profitability. We also provided guidelines for identifying market structure in practice. Among our conclusions are the following:

- Economic market structures can be grouped into four categories: perfect competition, monopolistic competition, oligopoly, and monopoly.
- The categories differ because of the following characteristics: The number of producers is many in perfect and monopolistic competition, few in oligopoly, and one in monopoly. The degree of product differentiation, the pricing power of the producer, the barriers to entry of new producers, and the level of nonprice competition (e.g., advertising) are all low in perfect competition, moderate in monopolistic competition, high in oligopoly, and generally highest in monopoly.
- A financial analyst must understand the characteristics of market structures in order to better forecast a firm's future profit stream.
- The optimal marginal revenue equals marginal cost. However, only in perfect competition does the marginal revenue equal price. In the remaining structures, price generally exceeds marginal revenue because a firm can sell more units only by reducing the per unit price.
- The quantity sold is highest in perfect competition. The price in perfect competition is usually lowest, but this depends on factors such as demand elasticity and increasing returns to scale (which may reduce the producer's marginal cost). Monopolists, oligopolists, and producers in monopolistic competition attempt to differentiate their products so that they can charge higher prices.
- Typically, monopolists sell a smaller quantity at a higher price. Investors may benefit from being shareholders of monopolistic firms that have large margins and substantial positive cash flows.
- Competitive firms do not earn economic profit. There will be a market compensation for the rental of capital and of management services, but the lack of pricing power implies that there will be no extra margins.

- While in the short run firms in any market structure can have economic profits, the more competitive a market is and the lower the barriers to entry, the faster the extra profits will fade. In the long run, new entrants shrink margins and push the least efficient firms out of the market.
- Oligopoly is characterized by the importance of strategic behavior. Firms can change the price, quantity, quality, and advertisement of the product to gain an advantage over their competitors. Several types of equilibrium (e.g., Nash, Cournot, kinked demand curve) may occur that affect the likelihood of each of the incumbents (and potential entrants in the long run) having economic profits. Price wars may be started to force weaker competitors to abandon the market.
- Measuring market power is complicated. Ideally, econometric estimates of the elasticity of demand and supply should be computed. However, because of the lack of reliable data and the fact that elasticity changes over time (so that past data may not apply to the current situation), regulators and economists often use simpler measures. The concentration ratio is simple, but the HHI, with little more computation required, often produces a better figure for decision making.

REFERENCES

Banker, R.D., I. Khosla, and K.K. Sinha. 1998. "Quality and Competition." *Management Science*, vol. 44, no. 9:1179–1192.

Chamberlin, Edward H. 1933. *The Theory of Monopolistic Competition*. Cambridge, MA: Harvard University Press.

Dorsey, Pat. 2004. *The Five Rules for Successful Stock Investing: Morningstar's Guide to Building Wealth and Winning in the Market*. Hoboken, NJ: John Wiley & Sons.

Friedman, Thomas L. 2006. *The World Is Flat: A Brief History of the Twenty-first Century*. New York: Farrar, Straus and Giroux.

Fudenberg, Drew, and Jean Tirole. 1984. "The Fat Cat Effect, the Puppy Dog Ploy and the Lean and Hungry Look." *American Economic Review*, vol. 74, no. 2:361–366.

Gómez-Ibáñez, José A. 2006. *Regulating Infrastructure: Monopoly, Contracts, and Discretion*. Cambridge, MA: Harvard University Press.

Kelly, Anthony. 2011. *Decision Making Using Game Theory: An Introduction for Managers*. Cambridge, UK: Cambridge University Press.

Krugman, Paul R. 1989. "Industrial Organization and International Trade." In *Handbook of Industrial Organization*, vol. 2. Edited by Richard Schmalensee and Robert Willig. Amsterdam: Elsevier B.V.

McCloskey, Donald. 1985. *The Applied Theory of Price*. 2nd ed. New York: Macmillan.

McGuigan, James R., R. Charles Moyer, and Frederick H. Harris. 2016. *Managerial Economics: Applications, Strategy and Tactics*. 14th ed. Mason, OH: Thomson South-Western.

Porter, Michael E. 2008. "The Five Competitive Forces that Shape Strategy." *Harvard Business Review*, vol. 86, no. 1:78–93.

Nicholson, Walter, and Christopher M. Snyder. 2016. *Microeconomic Theory: Basic Principles and Extensions*. 12th ed. Mason, OH: Thomson South-Western.

Schumpeter, Joseph A. 1942. *Capitalism, Socialism and Democracy*. New York: HarperCollins.

von Stackelberg, Heinrich. 1952. *The Theory of the Market Economy*. New York: Oxford University Press.

PRACTICE PROBLEMS

1 A market structure characterized by many sellers with each having some pricing power and product differentiation is *best* described as:

A oligopoly.

B perfect competition.

C monopolistic competition.

2 A market structure with relatively few sellers of a homogeneous or standardized product is *best* described as:

A oligopoly.

B monopoly.

C perfect competition.

3 Market competitors are *least likely* to use advertising as a tool of differentiation in an industry structure identified as:

A monopoly.

B perfect competition.

C monopolistic competition.

4 Upsilon Natural Gas, Inc. is a monopoly enjoying very high barriers to entry. Its marginal cost is $40 and its average cost is $70. A recent market study has determined the price elasticity of demand is 1.5. The company will *most likely* set its price at:

A $40.

B $70.

C $120.

5 The demand schedule in a perfectly competitive market is given by $P = 93 - 1.5Q$ (for $Q \leq 62$) and the long-run cost structure of each company is:

Total cost:	$256 + 2Q + 4Q^2$
Average cost:	$256/Q + 2 + 4Q$
Marginal cost:	$2 + 8Q$

New companies will enter the market at any price greater than:

A 8.

B 66.

C 81.

6 Companies *most likely* have a well-defined supply function when the market structure is:

A oligopoly.

B perfect competition.

C monopolistic competition.

7 Aquarius, Inc. is the dominant company and the price leader in its market. One of the other companies in the market attempts to gain market share by undercutting the price set by Aquarius. The market share of Aquarius will *most likely*:

A increase.

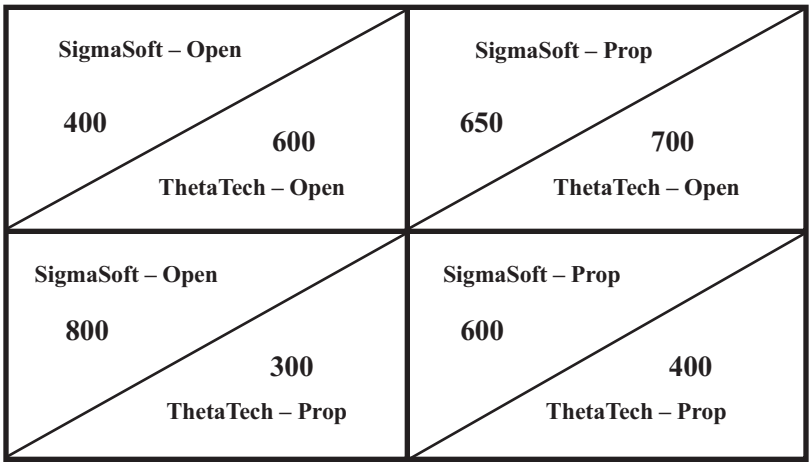
B decrease.

C stay the same.

8 SigmaSoft and ThetaTech are the dominant makers of computer system software. The market has two components: a large mass-market component in which demand is price sensitive, and a smaller performance-oriented component in which demand is much less price sensitive. SigmaSoft's product is considered to be technically superior. Each company can choose one of two strategies:

- *Open architecture (Open):* Mass market focus allowing other software venders to develop products for its platform.
- *Proprietary (Prop):* Allow only its own software applications to run on its platform.

Depending upon the strategy each company selects, their profits would be:



The Nash equilibrium for these companies is:

A proprietary for SigmaSoft and proprietary for ThetaTech.

B open architecture for SigmaSoft and proprietary for ThetaTech.

C proprietary for SigmaSoft and open architecture for ThetaTech.

9 A company doing business in a monopolistically competitive market will *most likely* maximize profits when its output quantity is set such that:

A average cost is minimized.

B marginal revenue equals average cost.

C marginal revenue equals marginal cost.

10 Oligopolistic pricing strategy *most likely* results in a demand curve that is:

A kinked.

B vertical.

C horizontal.

11 Collusion is *less likely* in a market when:

A the product is homogeneous.

B companies have similar market shares.

C the cost structures of companies are similar.

12 If companies earn economic profits in a perfectly competitive market, over the long run the supply curve will *most likely*:

A shift to the left.

B shift to the right.

C remain unchanged.

13 Over time, the market share of the dominant company in an oligopolistic market will *most likely*:

A increase.

B decrease.

C remain the same.

14 A government entity that regulates an authorized monopoly will *most likely* base regulated prices on:

A marginal cost.

B long run average cost.

C first degree price discrimination.

15 An analyst gathers the following market share data for an industry:

Company	Sales (in millions of €)
ABC	300
Brown	250
Coral	200
Delta	150
Erie	100
All others	50

The industry's four-company concentration ratio is *closest* to:

A 71%.

B 86%.

C 95%.

16 An analyst gathered the following market share data for an industry comprised of five companies:

Company	Market Share (%)
Zeta	35
Yusef	25
Xenon	20
Waters	10
Vlastos	10

The industry's three-firm Herfindahl–Hirschmann Index is *closest* to:

A 0.185.

B 0.225.

C 0.235.

17 One disadvantage of the Herfindahl–Hirschmann Index is that the index:

A is difficult to compute.

B fails to reflect low barriers to entry.

C fails to reflect the effect of mergers in the industry.

18 In an industry comprised of three companies, which are small-scale manufacturers of an easily replicable product unprotected by brand recognition or patents, the *most* representative model of company behavior is:

A oligopoly.

B perfect competition.

C monopolistic competition.

19 Deep River Manufacturing is one of many companies in an industry that make a food product. Deep River units are identical up to the point they are labeled. Deep River produces its labeled brand, which sells for $2.20 per unit, and "house brands" for seven different grocery chains which sell for $2.00 per unit. Each grocery chain sells both the Deep River brand and its house brand. The *best* characterization of Deep River's market is:

A oligopoly.

B perfect competition.

C monopolistic competition.

SOLUTIONS

1 C is correct. Monopolistic competition is characterized by many sellers, differentiated products, and some pricing power.

2 A is correct. Few sellers of a homogeneous or standardized product characterizes an oligopoly.

3 B is correct. The product produced in a perfectly competitive market cannot be differentiated by advertising or any other means.

4 C is correct. Profits are maximized when MR = MC. For a monopoly, MR = $P[1 - 1/E_p]$. Setting this equal to MC and solving for P:

$$\$40 = P[1 - (1/1.5)] = P \times 0.333$$
$$P = \$120$$

5 B is correct. The long-run competitive equilibrium occurs where MC = AC = P for each company. Equating MC and AC implies $2 + 8Q = 256/Q + 2 + 4Q$.

Solving for Q gives $Q = 8$. Equating MC with price gives $P = 2 + 8Q = 66$. Any price above 66 yields an economic profit because P = MC > AC, so new companies will enter the market.

6 B is correct. A company in a perfectly competitive market must accept whatever price the market dictates. The marginal cost schedule of a company in a perfectly competitive market determines its supply function.

7 A is correct. As prices decrease, smaller companies will leave the market rather than sell below cost. The market share of Aquarius, the price leader, will increase.

8 C is correct. In the Nash model, each company considers the other's reaction in selecting its strategy. In equilibrium, neither company has an incentive to change its strategy. ThetaTech is better off with open architecture regardless of what SigmaSoft decides. Given this choice, SigmaSoft is better off with a proprietary platform. Neither company will change its decision unilaterally.

9 C is correct. The profit maximizing choice is the level of output where marginal revenue equals marginal cost.

10 A is correct. The oligopolist faces two different demand structures, one for price increases and another for price decreases. Competitors will lower prices to match a price reduction, but will not match a price increase. The result is a kinked demand curve.

11 B is correct. When companies have similar market shares, competitive forces tend to outweigh the benefits of collusion.

12 B is correct. The economic profit will attract new entrants to the market and encourage existing companies to expand capacity.

13 B is correct. The dominant company's market share tends to decrease as profits attract entry by other companies.

14 B is correct. This allows the investors to receive a normal return for the risk they are taking in the market.

15 B is correct. The top four companies in the industry comprise 86 percent of industry sales: (300 + 250 + 200 + 150)/(300 + 250 + 200 + 150 + 100 + 50) = 900/1050 = 86%.

16 B is correct. The three-firm Herfindahl–Hirschmann Index is $0.35^2 + 0.25^2 + 0.20^2 = 0.225$.

17 B is correct. The Herfindahl–Hirschmann Index does not reflect low barriers to entry that may restrict the market power of companies currently in the market.

18 B is correct. The credible threat of entry holds down prices and multiple incumbents are offering undifferentiated products.

19 C is correct. There are many competitors in the market, but some product differentiation exists, as the price differential between Deep River's brand and the house brands indicates.